

```
Query Match: 18.17% Indels: 24
DB: 9 Gaps: 5
US-09-921-992-50 (1-372) x US-09-921-992-22 (1-670)
QY 133 ArgIleGlyValAsnAlaGlySerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152
DB 15 AGGATCGCGGTTAACCATGGTCTCTCTCCGAAAGAATGC-----TTTACCTACTTTT 68
QY 152 oThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172
DB 69 TACCCCGAGGGATGGTGAATGCCTCGGATTCATCAAAATTTGTAGCTTTAGA 128
QY 172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192
DB 129 TTTCCGCAACCTAGTCTGTTTCCATGAAGCGTCCCGGTACCGGTATGTTGCCGCTA 188
QY 192 rArgLeuLeuAlaLysGlnIleAspGln-----ProLeuHisLeuGlyI 207
DB 189 TCGCTCATGGTGAACAGTATGACGAGTTGGGCATGGATTATCCCTCCATCTAGGGGT 248
QY 207 eThrGluAlaGlyClyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuLeuLe 227
DB 249 TACCGAAGCGGGATGGGAATATGCGCGCATTAATCCACCGCTGGCATTCGCCACCT 308
QY 227 uLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluG 247
DB 309 TTTAGCTGATGGCATTTGGCGATATATCCGGGTATCCCTCACCAGAGCCCCGAAAGA 368
QY 247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267
DB 369 AATTCCCGTTGTACAGCATTTCCAGGCGTGGTTCGCGAAACCATGTTGGAATA 428
QY 267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLe 287
DB 429 TGTGGCTGTCTCTCTGTCGCGCACGTTGTTCAAC-----TT 467
QY 287 uGluGlnArgLeuGluAspIle-----IleThrProMetAsp-ValS 301
DB 468 GGAAGACGTGTACATGAAGTCCGAGATGCCACTAAACATCTAACCTGTTTAGCTTCG 527
QY 301 erIleLeGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValt 321
DB 528 NCGTCATGTGCTATTGTCAATGGCCCGGTGCAATGGCCGATGCGGCTATG 587
QY 321 hrGly 322
DB 588 TGGGT 592
RESULT 15
US-09-921-992-23
; Sequence 23, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohrer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavarapu
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT FILING DATE: 2001-08-06
; PRIOR FILING DATE: 2000-08-06
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 23
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Zea mays
;
FEATURE:
; NAME/KEY: unsure
; LOCATION: (1..596)
; OTHER INFORMATION: unsure at all n locations
US-09-921-992-23
Alignment Scores: 1.71e-30 Length: 596
Pred. No.: 326.00 Matches: 78
Score: 58.62% Conservatives: 41
Percent Similarity: 38.42% Mismatches: 62
Best Local Similarity: 17.47% Indels: 22
Query Match: 9 Gaps: 4
US-09-921-992-50 (1-372) x US-09-921-992-23 (1-596)
QY 133 ArgIleGlyValAsnAlaGlySerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152
DB 15 AGGATCGCGGTTAACCATGGTCTCTCTCCGAAAGAATGC-----TTTACCTACTTTT 68
QY 152 oThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172
DB 69 TACCCCGAGGGATGGTGAATGCCTCGGATTCATCAAAATTTGTAGCTTTAGA 128
QY 172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192
DB 129 TTTCCGCAACCTAGTCTGTTTCCATGAAGCGTCCCGGTACCGGTATGTTGCCGCTA 188
QY 192 rArgLeuLeuAlaLysGlnIleAspGln-----ProLeuHisLeuGlyI 207
DB 189 TCGCTCATGGTGAACAGTATGACGAGTTGGGCATGGATTATCCCTCCATCTAGGGGT 248
QY 207 eThrGluAlaGlyClyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuLeuLe 227
DB 249 TACCGAAGCGGGATGGGAATATGCGCGCATTAATCCACCGCTGGCATTCGCCACCT 308
QY 227 uLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluG 247
DB 309 TTTAGCTGATGGCATTTGGCGATATATCCGGGTATCCCTCACCAGAGCCCCGAAAGA 368
QY 247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267
DB 369 AATTCCCGTTGTACAGCATTTCCAGGCGTGGTTCGCGAAACCATGTTGGAATA 428
QY 267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLe 287
DB 429 TGTGGCTGTCTCTCTGTCGCGCACGTTGTTCAAC-----TT 467
QY 287 uGluGlnArgLeuGluAspIle-----IleThrProMetAsp-ValS 301
DB 468 GGAAGACGTGTACATGAAGTCCGAGATGCCACTAAACATCTAACCTGTTTAGCTTCG 527
QY 301 erIleLeGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValt 321
DB 528 NCGTCATGTGCTATTGTCAATGGCCCGGTGCAATGGCCGATGCGGCTATG 587
QY 321 hrGly 322
DB 588 TGGGT 592
Search completed: November 23, 2003, 21:59:12
Job time : 4946 secs
```

**THIS PAGE BLANK (USPTO)**

QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMetValValAsp 123  
Db 700 AAGATCGTGTCAACCCAGGAATTTGCGACAGCGGCCCGCCAGTTTGAGACATAGAT 759  
QY 124 CysAlaArgAspLys----- 128  
Db 760 TATACAGAAGATGAATATATCAGAAAGAACTCCAGCATATCGAGCAGGTCTTCACTCCTTTG 819  
QY 129 -----AsnIleProIleArgIleGlyValAsnAlaGlySerLeu 141  
Db 820 GTTGAGAAATGCAAAAGTACGGAGAGCAATCGGTATTGGGCAAAATCATGGAAGTCTT 879  
QY 142 GluLysAspLeuGlnGluLysTyrGlyProThrProGlnAlaLeuLeuGluSerAla 161  
Db 880 TCTGACCGTATCATGACGTATTACGGGAT--TCTCCCGAGGAATGTTGAATCTGCG 936  
QY 162 MetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLys 181  
Db 937 TTTGAGTTTGCAGAAATATGTCGGAATATTAGACTATCACAACTTTGTTTCTCAATGAAA 996  
QY 182 AlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIle----- 199  
Db 997 GCGAGCAACCCAGTATGTCGAGCGGTACCGTTTACTTGTGCTGAGATGTATGTT 1056  
QY 200 -----AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGly 216  
Db 1057 CATGGATGGATATCTTTGTCATTTGGGAGTTACTGAGCAGAGAGCGGAGATGGA 1116  
QY 217 AlaValLysSerAlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeu 236  
Db 1117 CGGATGAAATCTCGATTTGAATTTGGACGCTTCTTCAGACGGGCTCGGTGACACAATA 1176  
QY 237 ArgValSerLeuAlaAlaAspProValGluGluIle 248  
Db 1177 AGAGTTTCACTGACGGAGCCACCAGAGAGGAGATA 1212

## RESULT 13

US-10-066-543-1184  
; Sequence 1184, Application US/10066543  
; Publication No. US20030087818A1

## GENERAL INFORMATION:

; APPLICANT: Jiang, Yuqiu  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Indrias, Carol Yoseph  
; APPLICANT: Lodes, Michael J.  
; APPLICANT: Secrist, Heather  
; APPLICANT: Carter, Darrick  
; APPLICANT: Fanger, Gary R.  
; APPLICANT: Smith, Carole L.  
; APPLICANT: Durham, Margarita  
; APPLICANT: Stolk, John A.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.563  
; CURRENT APPLICATION NUMBER: US/10/066,543  
; CURRENT FILING DATE: 2002-01-31  
; NUMBER OF SEQ ID NOS: 3417  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1184  
; LENGTH: 706  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 2, 645, 659  
; OTHER INFORMATION: n = A,T,C or G

Alignment Scores: 5.99e-46 Length: 706  
Pred. No.: 452.00 Matches: 89  
Score: 73.29% Conservative: 29  
Percent Similarity:

Best Local Similarity: 55.28% Mismatches: 43  
Query Match: 24.22% Indels: 0  
DB: 14 Gaps: 0  
US-09-921-992-50 (1-372) x US-10-066-543-1184 (1-706)  
QY 10 ArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIleAla 29  
Db 218 AAAGAGAGTCGAGAAATACAAATCGGATGTTAAGATTGGGAAAGAAATCCCGTTGTG 277  
QY 30 ValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIleLys 49  
Db 278 ATTCAATCGATGATTAAACACGGAATCTCGGATGTAGAACGCGAGTGTACAGCAAAATTTTG 337  
QY 50 AlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAlaAla 69  
Db 338 GATTTTGAAGACCGCGTGTGAATTTGTTCTGTATGACGATCAATACAAAGAGCGCGCT 397  
QY 70 GluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHisPhe 89  
Db 398 ATGGCAATTTCCCGCGATTAAAGAGAGTTTCATATTTCTTTGTTAGCGGATATTCAATTT 457  
QY 90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109  
Db 458 GATTATCGACTTGCCTTATTGCGGATAGAACAGGGAATTGACAAATTTACGGATTATCC 517  
QY 110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129  
Db 518 GGAATATATCGTTCGAGAAATATTCGTTTGGTTCGGAAGCGGCAAAAGGAAGAGA 577  
QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlyLysTyr 149  
Db 578 ATTCGATTCGATTGAGTCAATGCAGGTTCTTTGGAAAAACATATCTTGGAAAAATAT 637  
QY 150 GlyGluProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAsp 169  
Db 638 GGAGCGGNAACGCGAGATGCTNTGTCAAAGCGCTATGATCATGTATAAACTGTTGGAA 697  
QY 170 Arg 170  
Db 698 CAA 700

## RESULT 14

US-09-921-992-22  
; Sequence 22, Application US/09921992  
; Patent No. US20020069426A1

## GENERAL INFORMATION:

; APPLICANT: Boronath, Albert;  
; APPLICANT: Campos, Narciso;  
; APPLICANT: Rodriguez-Concepcion, Manuel;  
; APPLICANT: Rohmer, Michel;  
; APPLICANT: Seeman, Myriam;  
; APPLICANT: Valentin, Henry E.;  
; APPLICANT: Venkatesh, Tyamagondlu V.;  
; APPLICANT: Venkatramesh, Mylavaram;  
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
; FILE REFERENCE: 16516.107/35-21(51897)US  
; CURRENT APPLICATION NUMBER: US/09/921,992  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/223,483  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 85  
; SEQ ID NO 22  
; LENGTH: 670  
; TYPE: DNA  
; ORGANISM: Zea mays  
US-09-921-992-22

Alignment Scores: 5.06e-32 Length: 670  
Pred. No.: 339.00 Matches: 81  
Score: 59.81% Conservative: 44  
Percent Similarity: 38.76% Mismatches: 60  
Best Local Similarity:









```
Db 303 TTCCATTATAAATTCGCTCTC---ATTGCGCGCTCAAAAGCGTGGATCGCATCAGGATTAAC 359
Qy 109 ProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db 360 CCCGAAACATCGCTCTTAAGAGAAGATCAAAAGCGTGGTGTATCTTGTAAAGAAAAA 419
Qy 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db 420 AACATTCTTAAGATTGGCGTGAATGCTGGGAGTTTAGAAAAGCAGTTTGTATCAAAA 479
Qy 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 480 TACGGA---CCCAACCCAAAGCGCTTAGAAAAGCGCTTTGTATAACGCCAAACTTTTA 536
Qy 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 537 GAAGATTGGATTTCACCAATTTAAGATTCTTTAAAGCGAGCGATGATTTCACACC 596
Qy 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 597 ATAGAAGCTTACAGGATGCTTCGCCCTCTTGTGATCTATCTTTCATTGGGGGTTACG 656
Qy 209 GluAlaGlyClyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 657 GAGCGGGGAATCTTTTAGCTCCAGTATCAAAATCCGCTATGCTTTAGGGGGCTTTTA 716
Qy 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluIle 248
Db 717 ATGAGGGGATGGGATAGATGAGTGGCGGTATCTCATCAGGGGAATTAGAAATGAATC 776
Qy 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 777 AAGTGGCCAGACCAATTTTACCCATAGCGCGGCTTGAAGCAATTAGTGGATTAATTGAT 836
Qy 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 837 TCTTGGCCCACTTCGCGGCGCATTAAGCCAAATTAGTGATATGCGCATCAAGGTAGAA 896
Qy 289 GlnArgLeuGluAspIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 897 AAACGCTTAAGCCACATCAAAACCCCTTTAGACATTAGCGTGATGGGTTCGGTGGTAA 956
Qy 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysSerGly 328
Db 957 GCTTTGGGTGAAGCAAGCATGACATGCGCATCGCTTTTGGGAATCGCAGCGGTTTG 1016
Qy 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
Db 1017 ATCATTAAAGAGGGTAAAGTCATTCAAAACTGGCTGAAAAGGATTATTTTGAAACTTTT 1076
Qy 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
Db 1077 GTGATGAAGTGGAAAATTTAGCTTAAGAAAAGAGAAAAAAGT 1118
```

## RESULT 8

```
US-10-156-761-1636
; Sequence 1636, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: HATTORI, YOSHIYUKI
; APPLICANT: SAKAKI, MASAHIRO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
```

```
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1636
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1155)
US-10-156-761-1636
```

```
Alignment Scores:
Pred. No.: 5,93e-87 Length: 1155
Score: 786.50 Matches: 164
Percent Similarity: 63.69% Conservative: 71
Best Local Similarity: 44.44% Mismatches: 121
Query Match: 42.15% Indels: 13
DB: 14 Gaps: 4
```

US-09-921-992-50 (1-372) x US-10-156-761-1636 (1-1155)

```
Qy 6 Profile---GlnArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAsp 24
Db 43 CCGATCGCGAAGCGGTGTCTCCGGCAGATCCAGTCCGACCGCTGGCGGTGGCGGC 102
Qy 25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThr 44
Db 103 ACGGCCCCGTGTCGTGTCAGTCGATGACGACGCGCTACGTTCGACATCGCGCACG 162
Qy 45 ValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64
Db 163 CTCAGCAGATCGCGAACTCACCGCTCGCGCTGCCAGATCGTCGCGCTCGCTCCCC 222
Qy 65 ThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuVal 84
Db 223 ACGAGGACGACGCGGACGCCCTCGCGGTATCGCCGCAAGTCGAGATCCCGTCGTC 282
Qy 85 AlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCys 104
Db 283 GCGGACATCCACTTCAGCGCGAAGTACGTGTTCCGCGCATCGAGCGGGCTGCGCGCG 342
Qy 105 LeuArgIleAsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAsp 123
Db 343 GTCCGGGTGAACCCGGGCAACATCAAGCAGTTCGACACAAGTCAAGGAGATCGCGCGC 402
Qy 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
Db 403 GCGGCAAGGAGACGCGCAGTCGATCCGCGTCAAGCGGCTCAAGCGGCTCGCTCGACCG 462
Qy 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
Db 463 CGGCTGCTGGAGAGTACGCGCAAGGCCACCCCGAGCGCTCGTCGAGTCGGCTGTGG 522
Qy 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
Db 523 GAGGCTGCTCTTCGAGGAGCAGCACTTCCGGGACATCAAGATCTCGGTCAAGCAAC 582
Qy 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
Db 583 GACCGGCTGTCATGTCGTCACCGGTACCGCGCAGCTGGCGCGCGAGTCGACATCCCGCTG 642
Qy 204 HisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
Db 643 CACCTCGGCTGACCGAGCGCGGCCCTTCCAGGCGACGATCAAGTCGCGCGCTCGCC 702
Qy 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAsp 243
Db 703 TTCGGCGCGCTGCTCAGCGAGGCGATCGCGCACACGATCCGCTCGCTCGACGCGCC 762
Qy 244 ProValGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
Db 763 CCGTTCGAGGAGATCAAGTTCGCGCATCCAGATCTCTGGAGTCTGGGCTTCAGGACGG 822
```

OTHER INFORMATION: a, t, c, g, other or unknown  
US-10-156-761-1

Alignment Scores:  
Pred. No.: 1,11e-83 Length: 9025608  
Score: 808.00 Matches: 164  
Percent Similarity: 64.11% Conservative: 70  
Best Local Similarity: 44.93% Mismatches: 119  
Query Match: 43.30% Indels: 12  
DB: 14 Gaps: 3

US-09-921-992-50 (1-372) x US-10-156-761-1 (1-9025608)

QY 8 GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro 27  
Db 3141566 GAGCGCGGAAGAGCGCGCAGATCCAGTCCGAACCGTGGCGGAGAGCGACCC 3141507  
QY 28 IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGln 47  
Db 3141506 GTCTCGGTGCAGTCGATGACGACGACGCGTACGTCGACATCGCGCCACGCTCCAGCAG 3141447  
QY 48 IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67  
Db 3141446 ATCGCCGAGCTCAGCGGCTCCGCTGCGCAGATCGTCGCTGCGCTGCCCGCCAGCGAGAC 3141387  
QY 68 AlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIle 87  
Db 3141386 GACCGCGAGCCCTCGCGGTATCGCCGCAAGTCGAGATCCCGCTGATCGCGGACATC 3141327  
QY 88 HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIle 107  
Db 3141326 CACTTCAGCGGAAGTACGTGTTCCGCGGATCGCGCGGCTGCGCGGCGTCCGGTG 3141267  
QY 108 AsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArg 126  
Db 3141266 AACCCGGGCAACATCAAGCATGTTCCGACGACAGGTCAAGGAGATCCCAAGCGCGCAAG 3141207  
QY 127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146  
Db 3141206 GAGCAGCGCAGCGCGATCCGATCGCGGTCAAGCGGGCTCGTCCACCGCGGCTGCTG 3141147  
QY 147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166  
Db 3141146 GAGAGTACGGCAAGCAAGCCCGAGGCGCTCGTCAGTCGCGCTGTGGAGGCGTCTG 3141087  
QY 167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186  
Db 3141086 CTCCTCGAGGAGCAGCACTTCGCGGACATCAAGATCTCGGTCAAGCACAACGACCGGTG 3141027  
QY 187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206  
Db 3141026 GTCATGCTCAACCGGTACCGCCAGCTGGCGGCGCAGTACCGCTGACCTCGGCTCGGC 3140967  
QY 207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
Db 3140966 GTGACCGAGGCGCGCCGCTTCAGGCGCAGCATCAAGTCGCGCTCGCTTCGCGCGG 3140907  
QY 227 LeuLeuSerGluLysIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246  
Db 3140906 CTGCTCAGCGAGGGCATCGCGCAGCATCCCGGTCTCGTGAAGCGCGCCCGCTCGAG 3140847  
QY 247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266  
Db 3140846 GAGATCAAGTTCGGCATCCAGATCTCTGGAGTCTGAACTCGCGCAGCGCGCTGGAG 3140787  
QY 267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286  
Db 3140786 ATCGTCTCTCGCGCTCGCGGAGCGCGCCAGGTGACGTCTACAAGCTGGCGGAGGAG 3140727  
QY 287 LeuGluGlnArgLeuAlaAspIleIleThrProMetAspValSerIleGlyCysVal 306  
Db 3140726 GTCACGGCGGCTCGGAGGCGATGGAGGTCGCCCTCGCGGTCTACGCTGCGGTGCGTC 3140667

QY 307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326  
Db 3140666 GTCAACGGCGCGGGGAGGCCCGGAGCGACCTCGCGCTCGCTCCGCAACGCGCAAG 3140607  
QY 327 SerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAsp 346  
Db 3140606 GGACAGATCTCTGTAAGGGCGGAAATCATCAAGACGCTGCCCGAGTCTGAGGAG 3140547  
QY 347 GlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAsp 366  
Db 3140546 ACCCTG-----ATCAGAGGCGGATGAAGATC--- 3140520  
QY 367 ValGlnGlnValGlu 371  
Db 3140519 GCCGAGCAGATGGAG 3140505  
RESULT 7  
US-09-881-752A-195  
; Sequence 195, Application US/09881752A  
; Patent No. US20020115078A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleanthous, Harold  
; APPLICANT: Al-Garawi, Amal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Oomen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 06132/041002  
; CURRENT APPLICATION NUMBER: US/09/881,752A  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 08/833,457  
; PRIOR FILING DATE: 1997-04-01  
; NUMBER OF SEQ ID NOS: 370  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 195  
; LENGTH: 1180  
; TYPE: DNA  
; ORGANISM: Helicobacter pylori  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (51)...(1127)  
US-09-881-752A-195  
Alignment Scores:  
Pred. No.: 7,26e-88 Length: 1180  
Score: 794.00 Matches: 162  
Percent Similarity: 66.38% Conservative: 73  
Best Local Similarity: 45.78% Mismatches: 117  
Query Match: 42.55% Indels: 2  
DB: 10 Gaps: 2  
US-09-921-992-50 (1-372) x US-09-881-752A-195 (1-1180)  
QY 9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28  
Db 63 AGAGTTAAGACCAAGCAAAATTTTATCGGTGCGGTGGCCATAGGGGGTGTATCTCCCAT 122  
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48  
Db 123 AGCAGCGCAAGCATGACCTTTAGCAAAACCGCTGATATTGAAGCAGCTAAAAATCAAT 182  
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68  
Db 183 GACAGACTCAAACTCGCGGGCGCGATTAGTAGGGTGGCGGTGAGTAATGAAAAGGAC 242  
QY 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88  
Db 243 GCTCTAGCTTTAAAGAAATGAAAAAGTGTCCCTTTTGCCTTTTAAATCGCTGATTTTAT 302  
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108

924 GGTCGGGAGAGCGCGGAGCCGATATCCGGATCGCGGCACGCCG 974

## RESULT 5

US-10-156-761-2548

; Sequence 2548, Application US/10156761

; Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, S.

APPLICANT: IKEDA, HARUO

APPLICANT: IKEDA, HARUO  
APPLICANT: ISHIKAWA, JUN

APPLICANT: ISHIKAWA, YOUN  
APPLICANT: HORIKAWA, HIROSHI

APPLICANT: HOKINAWA, HIROSHI  
APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SHIBA, TADAYOSHI  
APPLICANT: SAKAKI YOSHIYUKI

APPLICANT: SAKAKI, YOSHIYUKI  
APPLICANT: HATTORI, MASAHIRA

APPLICANT: HATTORI, MASAHIRA  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 349-363

FILE REFERENCE: 249-262  
CURRENT IDENTIFICATION NUMBER: 118/10/155 751

; CURRENT APPLICATION NUMBER: US/10/156,761  
 ; CURRENT FILING DATE: 2003-05-29

; CURRENT FILING DATE: 2002-05-29  
 ; PRIOR APPLICATION NUMBER: JP 2001-204088

;  
;  
PRIOR APPLICATION NUMBER: JP  
PRIOR APPLICATION NUMBER: 2003-0530

; PRIOR FILING DATE: 2001-05-30

;  
;  
PRIOR APPLICATION NUMBER: JP

; PRIOR FILING DATE: 2001-08-

; NUMBER OF SEQ

; SEQ ID NO 2548

; LENGTH: 1155

TYPE: DNA

ORGANISM: *Staphylococcus aureus*

NAME/KEY: CDS

; LOCATION: (1)..(1155)

US-10-156-761-2548

Alignment Scores:

Alignment Scores:			
Pred. No.:	1.31e-89	Length:	1155
Score:	808.00	Matches:	164
Percent Similarity:	64.11%	Conservative:	70
Best Local Similarity:	44.93%	Mismatches:	119
Query Match:	43.30%	Indels:	12
DB:	14	Gaps:	3

Score:	808.00	Matches:	164
--------	--------	----------	-----

Percent Similarity: 64.11% Conservative: 70

Best Local Similarity: 44.93% Mismatches: 119

Query Match:	43.30%	Indels:	12
--------------	--------	---------	----

DB:	14	Gaps:	3
-----	----	-------	---

US-09-921-992-50 (1-372) x US-10-156-761-2548 (1-1155)

Qy		8	GlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro	27
Dd		49	GAGCGCGGAAGACGGCGGCATCAGGTTCGAAACCGTGCCGTGGCGGAGACGCCACC	108
Qy		28	IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGln	47
Dd		109	GTCCTCGGTGCAGTCGATGACGACGCGGTACTCTCGACATCGGCGCCACGCTCCAGCAG	168
Qy		48	IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp	67
Dd		169	ATCGCCGAGCTCACGCGCTCCGCTGCCAGATCGTCCGTGTGGCTGCCCCACCGCAGAC	228
Qy		68	AlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIle	87
Dd		229	GACCGGAGCCCTCCGGGTCTATCGCCCAAGTCGACAGTCCCCTGATCCGCGACATC	288
Qy		88	HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAlaSpCysLeuArgile	107
Dd		289	CAC TTC CAG CGG AAG TAC GTT GCG GCG CAT CGA GCG GCG GTG CGG CGT CCG GGT G	348
Qy		108	AsnProGlyAsnIleGlyAsn--GluGluArgIleArgMetValValAlaSpCysAlaArg	126
Dd		349	AACCCGGGCACATCAAGCAGTTGCACGACACAGGTTCAAAGGATCGCCAAGCGCGCGAAG	408
Qy		127	AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln	146
Dd		409	GACACGGCACGCCGATCCGCATCGCGCTCAACGGGGGCTCGCTCGACCGCGCGCTGCTG	468
Qy		147	GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValasp	166
Dd		469	GAGAAGTAGCGAGGCCACCCCGAGGCGCTCGTCGAGTCGGCTCGGGAGGCGCTCG	528



```
Alignment Scores:
Pred. No.: 5,01e-183 Length: 1830121
Score: 1603.50 Matches: 320
Percent Similarity: 93.66% Conservative: 20
Best Local Similarity: 88.15% Mismatches: 22
Query Match: 85.93% Indels: 1
DB: 14 Gaps: 1

US-09-921-992-50 (1-372) x US-10-329-960-1 (1-1830121)
QY 4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23
Db 390972 CAGCCAACTATTAAAGCGTCGTGAATCGACAAAATTTATGTGGAAATGTACCAATTTGGT 391031
QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43
Db 391032 GGGGATGCGGCTATTGCGGTCAATCAATGACAAATCTCGCACCTGATGTGGAAGCG 391091
QY 44 ThrValAsnGlnIleLysAlaLeuGluAtgValGlyAlaAspIleValArgValSerVal 63
Db 391092 ACAGTTGCTCAATTAATCATTTAGAACGTTGTTGGTCAGATATTGTTCTGTATCTGTT 391151
QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83
Db 391152 CCAACAATGATGTCGCGGAAGCATTTAAACAAATTTAAACAAAGTGAATGTTCCGCTC 391211
QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
Db 391212 GTAGCAGATATTCAATTTGCACTATCGTATCGCTTAAAGTCGAGAATATGAGTGGAT 391271
QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
Db 391272 TGTTTACGTATCAATCTGGCAACATTTGTCGTGAAGATCGCTCGTGGCTGTTGTTGAT 391331
QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
Db 391332 TGTGCGGAGACAAAATATTCGGATTTCGTATTGGTGAATGCAGGCTCTTTTAGAAAAA 391391
QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSerAlaMetArg 163
Db 391392 GATTTTCAAGAAAATATGCGAACCAACCGCAGAACGCTTGTAGAACCTCGCATCCGATCGCT 391451
QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
Db 391452 CATGTAGAATCTTAGATCGCTTAATCTTCACTCAGTTTAAAGTAGCGTAAAGCGCTCC 391511
QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
Db 391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTTAAACAGCGCTTTA 391571
QY 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
Db 391572 CATTTAGGCATTACAGAAGCAGGTGGCGCAGCGGCTGGTGCATGAATTCGCAAGTGGGT 391631
QY 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
Db 391632 TTAGGAATGTTATTAGCTGAGGCGATTGGCGATACACTACGCTCTCTTTGGCGCGGAGAT 391691
QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
Db 391692 CCTGTAGAGGAATCAAAAGTCGGTTTGATATTTTGAAATCTTTACGGATTCTGTTCAAGA 391751
QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
Db 391752 GGAATTAACCTTTATTGCTTGCCCAACCTGTTCTCGCAAGAAATTTGATGTAAATCGGTACA 391811
QY 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
Db 391812 GTAAATGCGCTAGAACACCCCTTGAAGATATTATTACCAATGGATGTATCTATTATC 391871
QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
Db 391871 GTAAATGCGCTAGAACACCCCTTGAAGATATTATTACCAATGGATGTATCTATTATC 391931

Db 391872 GGTGTGTGTAGTGAATGGTCTCGCGAGGCACTCGTCTCCGATCTCGGCGTAAACGGCGGT 391931
QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
Db 391932 AACAAAAAAGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTGTATAACGAA 391991
QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
Db 391992 GATATAGTGAACCAATTAGAACCAAAATTCGTGCGAAGTGCACAGACAGATCCAAAA 392051
QY 363 ArgArgIle 365
Db 392052 AACAGAAAT 392060

RESULT 3
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
; US-09-790-988-1

Alignment Scores:
Pred. No.: 5,36e-144 Length: 640681
Score: 1282.00 Matches: 242
Percent Similarity: 83.88% Conservative: 65
Best Local Similarity: 66.12% Mismatches: 57
Query Match: 68.70% Indels: 2
DB: 10 Gaps: 2

US-09-921-992-50 (1-372) x US-09-790-988-1 (1-640681)
QY 1 MetHisAsnGlnAlaPro---IleGlnArgArgLysSerThrArgIleTyrValGlyAsn 19
Db 314272 ATGAATAATAAGTGTAAATTTATCAATAGAGAAAATCTGATCGTATTATGTTGAAAA 314331
QY 20 ValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThr 39
Db 314332 GTCGCTATTGGCAATAATGCGCAATATCAGTTCAATCTATGACAAATCTCGTACTACT 314391
QY 40 AspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleVal 59
Db 314392 AATACTCTGAACATTTAATCAATCTTAGAGTTACAAAAAGTAGGAGTAGATATTGTT 314451
QY 60 ArgValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnInVal 79
Db 314452 CGTATTTCTATACCAATTTAAAGCTGCAGAAATCATTTCAAGAAATAAAAAACAACA 314511
QY 80 AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99
Db 314512 AATGTTCCATTGTCAGATATACATTTGATTGATACAGATTAGCTTTTCAAGCTATATAAA 314571
QY 100 TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArg 119
Db 314572 TATGTTGCAAGATTGTTTGAAGATTTAATCTCGGAATATTGGAATATAAAGAGATATCA 314631
QY 120 MetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGly 139
Db 314631 TATGTTGCAAGATTGTTTGAAGATTTAATCTCGGAATATTGGAATATAAAGAGATATCA 314691
```

```
;/ NAME/KEY: misc_feature
;/ LOCATION: (44416)..(44416)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44905)..(44905)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44975)..(44975)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (45593)..(45593)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (45732)..(45732)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (47036)..(47036)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51334)..(51334)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51602)..(51602)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51786)..(51786)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51805)..(51805)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (55369)..(55369)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (65309)..(65309)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (65313)..(65313)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (80024)..(80024)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (100091)..(100091)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (102696)..(102696)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (105121)..(105121)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (107248)..(107248)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
```

```
;/ LOCATION: (117136)..(117136)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (119750)..(119750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (119924)..(119924)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (120038)..(120038)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (121344)..(121344)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (122167)..(122167)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (122336)..(122336)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (131340)..(131340)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (131360)..(131360)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (139910)..(139910)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (140398)..(140398)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (142750)..(142750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145058)..(145058)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145171)..(145171)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145942)..(145942)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (147197)..(147197)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (150841)..(150841)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (152500)..(152500)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (152530)..(152530)
```



```
; NAME/KEY: CDS
; LOCATION: (1)..(1119)
US-09-921-992-3

Alignment Scores:
Pred. No.: 2,37e-220 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-921-992-50 (1-372) x US-09-921-992-3 (1-1119)
QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATTAACGAGGCTCCAAATCAACGTAGAAATCAACACAGTATTACGTTGGGAATGG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
DB 61 CCGATTGGCGATGGTCTCCCATCGCGTACAGTCCATGACCAATACGCGTAGCAGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGCGCTGATATCGTCCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80
DB 181 GTATCCGTACCGACAGTGGACGGCGAGAGCGCTTCAAACTCATCAAAACAGCAGGTAAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTGCCCGTGGTGGCTGACATCCACTTCGACTATCGCATTCGCGTGAAGTAGCGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
DB 301 GGGCTCGATTGCTGGGTATTAAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTTCGATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
DB 361 GTGGTTGACTGTCCGGCGGATAAAACATTCCTCGTATTCGGCTTAAACGGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
DB 421 CTGGAAAAGATCTGCAAGAAAGATATGGCAACCGCGCGAGCGGTGCTGGATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACCTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGTGTACGCTCTCCCTCGCTGTAGTCTTATCGTTGCTGGCAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCCGTTGCATCTGGGGATCACCGAAGCGCGTGGTGGCGGACGCGGGCAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGTTTAGTCTGCTGCTGTCTGAAGGCATCGGCACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGCATCCGGTCAAGAGATCAAAAGTCGGTTTCGATATTTGAAATCGCTGCGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCGGAGGATCAACTTCATCGCTGCCACCGCTGTTTCGGCTCAGGAATTGATGT 840
QY 281 IleGlyThrValAsnAlaLeuGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTACGGTTAACGGCTGGAGCAACGCTGGAAGATATCATCTCCGATGGACGTT 900
```

```
301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
901 TCGATTATCGGCTGGTGGTGAATGCCAGGTGAGCGCTGGTTCTTACATCTCGCGCTC 960
321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
961 ACCGGCGGCAACAAGAAAGCGCTCTATGAAGATGGCGTGGCGCAAGACCGTCTGGAC 1020
341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
1021 AACACGATATATCATCGACCGTGAAGCACCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080
361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
1081 GAAGCGCTCGAATTGACGTTGACGAGGTGGAATAA 1116

RESULT 2
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 23, 2003, 18:51:50 ; Search time 402 Seconds  
(without alignments)  
3024.309 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAIQRRKSTRIYGVN.....RAKASQLEARRIDVQVQEK 372

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09921992/runat\_21112003\_162548\_18522/app\_query.fasta\_1.519  
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09921992@cgn\_1\_1\_347@runat\_21112003\_162548\_18522  
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	1119	9	US-09-921-992-3
					Sequence 3, Appli

2	1603.5	85.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
3	1282	68.7	640681	10	US-09-790-988-1	Sequence 1, Appli
4	822	44.1	1083	10	US-09-974-300-1692	Sequence 1692, Ap
5	808	43.3	1155	14	US-10-156-761-2548	Sequence 2548, Ap
6	808	43.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
7	794	42.6	1180	10	US-09-881-752A-195	Sequence 195, App
8	786.5	42.1	1155	14	US-10-156-761-1636	Sequence 1636, Ap
9	761	40.8	1134	10	US-09-738-626-2205	Sequence 2205, Ap
10	751.5	40.3	3309400	10	US-09-738-626-1	Sequence 1, Appli
11	462	24.8	2520	10	US-09-712-363-110	Sequence 110, App
12	452	24.2	706	14	US-09-921-992-1	Sequence 1, Appli
13	439	24.2	670	14	US-10-066-543-1184	Sequence 1184, Ap
14	339	18.2	670	9	US-09-921-992-22	Sequence 22, Appl
15	326	17.5	596	9	US-09-921-992-23	Sequence 23, Appl
16	321	17.2	584	9	US-09-921-992-21	Sequence 21, Appl
17	292	15.6	33675	9	US-09-921-992-2	Sequence 2, Appli
18	277	14.8	705	9	US-09-921-992-34	Sequence 34, Appl
19	262	14.0	601	9	US-09-921-992-11	Sequence 11, Appl
20	259	13.9	528	9	US-09-921-992-15	Sequence 15, Appl
21	258	13.8	379	9	US-09-921-992-16	Sequence 16, Appl
22	243	13.0	511	9	US-09-921-992-36	Sequence 36, Appl
23	241	12.9	498	9	US-09-921-992-37	Sequence 37, Appl
24	230	12.3	403	9	US-09-921-992-24	Sequence 24, Appl
25	221	11.8	938	9	US-09-921-992-13	Sequence 13, Appl
26	218	11.7	395	9	US-09-921-992-18	Sequence 18, Appl
27	215.5	11.5	464	9	US-09-921-992-33	Sequence 33, Appl
28	214	11.5	432	9	US-09-921-992-14	Sequence 14, Appl
29	212	11.4	395	9	US-09-921-992-17	Sequence 17, Appl
30	210	11.3	564	9	US-09-921-992-35	Sequence 35, Appl
31	198.5	10.6	443	9	US-09-921-992-12	Sequence 12, Appl
32	192.5	10.3	293	9	US-09-921-992-25	Sequence 25, Appl
33	179	9.6	388	9	US-09-921-992-44	Sequence 44, Appl
34	174	9.3	440	9	US-09-921-992-38	Sequence 38, Appl
35	170.5	9.1	211	9	US-09-921-992-45	Sequence 45, Appl
36	166.5	8.9	456	9	US-09-921-992-26	Sequence 26, Appl
37	166	8.9	348	10	US-09-974-300-6102	Sequence 6102, Ap
38	157.5	8.4	869	9	US-09-921-992-42	Sequence 42, Appl
39	155.5	8.3	551	9	US-09-921-992-41	Sequence 41, Appl
40	155	8.3	430	9	US-09-921-992-29	Sequence 29, Appl
41	155	8.3	615	9	US-09-921-992-6	Sequence 6, Appli
42	154.5	8.3	619	9	US-09-921-992-27	Sequence 27, Appl
43	151	8.1	422	9	US-09-921-992-28	Sequence 28, Appl
44	150	8.0	589	9	US-09-921-992-7	Sequence 7, Appli
45	149.5	8.0	399	9	US-09-921-992-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1

US-09-921-992-3  
; Sequence 3, Application US/09921992  
; Patent No. US2002008946A1  
; GENERAL INFORMATION:  
; APPLICANT: Boronat, Albert;  
; APPLICANT: Campos, Narciso;  
; APPLICANT: Rodriguez-Concepcion, Manuel;  
; APPLICANT: Rohrer, Michel;  
; APPLICANT: Seeman, Myriam;  
; APPLICANT: Valentin, Henry E.;  
; APPLICANT: Venkatesh, Mylavara  
; APPLICANT: Venkatesh, Mylavara  
; TITLE OF INVENTION: Methy-D-Erythritol Phosphate Pathway Genes  
; FILE REFERENCE: 1616.107/35-21(51897)US  
; CURRENT APPLICATION NUMBER: US/09/921,992  
; PRIOR FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/223,483  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 85  
; SEQ ID NO 3  
; LENGTH: 1119  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:



QY 226 LeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProVal 245  
Db 694 GGGTTATTGATGAAGGCAITGGGATACGATCGTATTTCGTTGCTGTAACCTGAA 753  
QY 246 GluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIle 265  
Db 754 GATGAATCAAGATCGGTTTGTATCTTAAATCGCTTGGCTTACGCTTCTAACGGGTATC 813  
QY 266 AsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsn 285  
Db 814 AACTTATTGCTGTCCAAAGTTGTTCTCGCCAGAAATTTAAACGTGATTCAGGTGATCGAG 873  
QY 286 AlaLeuGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCys 305  
Db 874 GCTTTAGAGAGCGTTTAGAAGATATCCGTACCAATGACCGTTTCGGTTATTGGCTGT 933  
QY 306 ValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325  
Db 934 AAGGTAATGGCCAGGTGAAGCAAAAGAACGAGATATCGGGTTGTTGGGGCTGCGCCT 993  
QY 326 LysSerGlyLeuTyrgluAspGlyValArgLysAspArgLeuAspAsnAspMetIle 345  
Db 994 CGTTCATTGCTTATCGTAATGTTGAGAAAGCCATTTAATTGATACCAATCAATTTGGTT 1053  
QY 346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIle 365  
Db 1054 GATGAATCGAACTATGTTGCTCAACGTTGTTCAAGAGCTTGAAGAAGCTTAATCTAAA 1113  
QY 366 AspVal 367  
Db 1114 GAAATT 1119

RESULT 15

US-08-827-190-9/c

Sequence 9, Application US/08827190  
Patent No. 5858367  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 886 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-827-190-9

Alignment Scores:  
Pred. No.: 2 68e-84 Length: 886  
Score: 812.00 Matches: 158  
Percent Similarity: 70.65% Conservative: 49  
Best Local Similarity: 53.92% Mismatches: 86  
Query Match: 43.52% Indels: 0  
DB: 2 Gaps: 0

US-09-921-992-50 (1-372) x US-08-827-190-9 (1-886)

QY 30 ValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIleLys 49  
Db 881 ATCAAGCATGACAAACAACAAACACATGATGTAGAAGCAACGGTTGCGGAATTTAAC 822  
QY 50 AlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAlaAla 69  
Db 821 CGTTTGGCTGAAAGCCGGATGCCAAATCGTTCCGGTAGCATGTCCGGATGAACCGCGGCA 762  
QY 70 GluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHisPhe 89  
Db 761 AACGCCATTGCGGATATTAAAGCCGATTCATTCCTCTCGTTGTTGACATACATTTC 702  
QY 90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109  
Db 701 GATTATAAACTTGGCTTGAAGCCATTGAAGCGCGGCAGATAAAATCCGAATCAACCCC 642  
QY 110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129  
Db 641 GGCAATATCGCGCGCGCGGCGGAAAGTTGAAGCGGTTGTTAAAGCGGCCAAAGCAAGGC 582  
QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLysTyr 149  
Db 581 ATTCCGATCAGATCGAGTAAACCGCGGTTTATTGGAAAAACGGATTTTAGAAAAATAC 522  
QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169  
Db 521 GGTATTCGACTGCGGATGGTGTAGAAAGCGCATTCATCACATTTAAATTTCTTGAG 462  
QY 170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189  
Db 461 GATCTTGATTTTCCAGCATATTATTGTCAGCATGAAGCGCTCTGACGTGAACCTTGAATC 402  
QY 190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuIleThrGlu 209  
Db 401 GAGGCTTATGAAGAACGAGCGAAAGCGTTGACTACCGCTTCACCTCGGATCACCGAG 342  
QY 210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSer 229  
Db 341 TCAGGAACACTGTTTCCGGCACAGTAAAGCGCAGCAGGACTCGCGCCCATTTTAAGC 282  
QY 230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLys 249  
Db 281 AAAGGCATCGGGAACACCATTCGCGCATTTTCACTAAGCGCAGACCTCTGTAGAGAGGTAAA 222  
QY 250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269  
Db 221 GTAGCAAGGAGCTTCTGAATCTTTTCGGCTTAGCTTCAATGCTGCCACGCTCATCTCA 162  
QY 270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289  
Db 161 TGCCGACTTTCGGCGCGTATTGAGATTGATCTAATCAGCATTCGCAATGAAGTGAAGAG 102  
QY 290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309  
Db 101 TATATTTCTAAGATAAAAGCGCGGATTAAGTTGCTTCTCGGCTGCGCTGTAACCGCA 42  
QY 310 ProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322  
Db 41 CCTGTGAAGCGGAGAGAGCTGATATCGAATTCGCGGC 3

Search completed: November 23, 2003, 19:44:10  
Job time : 952 secs

TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-170-187-10

Alignment Scores:  
Pred. No.: 111e-146 Length: 886  
Score: 1353.00 Matches: 268  
Percent Similarity: 96.26% Conservative: 15  
Best Local Similarity: 91.16% Mismatches: 11  
Query Match: 72.51% Indels: 0  
DB: 4 Gaps: 0

US-09-921-992-50 (1-372) x US-09-170-187-10 (1-886)

```
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
DB 884 GCTGTCCTCATCTATGACGATATCGCCACGACGAGATGTTGAGCCCACTGTGGCGCAATTC 825
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 824 CAATCACTTGAGCGTGTAGGTGTATATCGTCCGCGTGTCTGTTCTCTACGATGATGCA 765
QY 69 AlaGluAlaPheLysLeuLysGlnValAsnValProLeuValAlaAspIleHis 88
DB 764 GCAGAAAGCCCTTTAAATTAATTAAGCAGCGCGTGAATGTGCCATTTGTTGCGGATATTCAC 705
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysValArgIleAsn 108
DB 704 TTTGACTACCGTATCCGATGAAAGTGGCTGAATATGTTGTTGACTGCTCAGCAATTAAC 645
QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 644 CCAGGTAATATCGCAGTGAAGAGCGTATTCCGCAAGTGTTCATAGTGTCTGTCATCAC 585
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGluLys 148
DB 584 AACATTTCTATCCGTATAGGGGTCAATATGGCGGTCTCATCGGAAAGATATCCCAAGAAAAA 525
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 524 TACGGTGAGCCCAACACTGAGGACATTTGGTTGAATCAGCAATGCGACATGTTGATATCTTG 465
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 464 GACAGCGTGAATTTTCGATCAGTTCAGGTCAAGTGTAAAGCGTCGATGCTTTCTTGCC 405
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 404 GTCGGCTCTTATCGTTTATTGGCGCAAAAAATTTGATCAACCACTTCACCTCGGTATTACA 345
QY 209 GluAlaGlyValAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 344 GAAGCGGTGGGCTCGTCTGTTTCTAGTGAATCAGCAATTCGTTGTTGTTGTTGTTG 285
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB 284 GCTGAAGTATCGGCATACGTTACGTTATCTACTCGCGCAGATCTCTGTTGAGGAAGTG 225
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 224 AAAGTCGGTTTGTATATCTTAAATTCGTTACCGATCGCTCAGCTGCGCATCACTTTATT 165
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 164 GCTTGCCCAACCTGTTTACGCCCAAGAAATTTGATGTTGTTGTTGTTGTTGTTGTTG 105
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 104 CAGCGCTCGAAGATATTATCAGCGCGATGGATGCTCTATTATTGTTGTTGTTGTTGTTG 45
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
DB 44 GGCCCGGGTGAAGCCGAGGTTTCTACTTTTAGGTGTGGCTGGC 3
```

## RESULT 14

US-09-328-352-3780  
Sequence 3780, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Breton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
FILE REFERENCE: GTC99-03PA  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 3780  
LENGTH: 1137  
TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-3780

Alignment Scores:  
Pred. No.: 2.67e-128 Length: 1137  
Score: 1195.00 Matches: 231  
Percent Similarity: 81.77% Conservative: 65  
Best Local Similarity: 63.81% Mismatches: 66  
Query Match: 64.04% Indels: 0  
DB: 4 Gaps: 0

US-09-921-992-50 (1-372) x US-09-328-352-3780 (1-1137)

```
QY 6 ProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGly 25
DB 34 CCAATTAACCGTCGACCAACACGTAATCCGTGTTGGTTCGTCTATGTCGGTGGCGAT 93
QY 26 AlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrVal 45
DB 94 GCACCTATTAGTGTGCAAGATGACAAATACCGAACTTGGCATGTTGCGCACTGTG 153
QY 46 AsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThr 65
DB 154 GCTCAGATTGACGGTGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 213
QY 66 MetAspAlaAlaGluAlaPheLysLeuLysGlnValAsnValProLeuValAla 85
DB 214 ATGGAAGTGTCTGAAGCCCTTTGGTCAATCCGTAAGCGTGTTCATGTTCCATTTAGTAGT 273
QY 86 AspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeu 105
DB 274 GATATCCATTGTGACCATAGAAATTCCTTGGCAGTTGACGATTATGTCGGACTGCTTG 333
QY 106 ArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAla 125
DB 334 CGTATTAAACCCGGCAATATCGTTTCAGACCAAGAAAGTTTCGTGAAGTTTCGTGGCTGGCA 393
QY 126 ArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeu 145
DB 394 CGTCATCAGCGTATTTCATGCGTATGTCGTAATTCGCGGTCTCTAGAAAAAGATTTA 453
QY 146 GlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisVal 165
DB 454 CAGAAAAATATGGCAGCTTACAGGCGAGGCACTTCTTCAGTCAGCTTTCAGTCATATT 513
QY 166 AspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspVal 185
DB 514 GATATTTAGACCGTCTTGACTTCCATGAGTTTAAAGTCAGTGAAGATCAAAATGTG 573
QY 186 PheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeu 205
DB 574 TTTTAAACCATGGATGCTTATGCTTCTCTCAACAAATGATAATCCATACACCTT 633
QY 206 GlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGly 225
DB 634 GGAGTGTGAAGCTGGTATTATTCAGGTACTGTGAAATTCAGGATTCGTTCTTGTGT 693
```

APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:

QY 201 GlnProLeuHisLeuGlyLeuThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220  
Db 1527 CAGCCCTGCACCTGGGCATACCCAGGCGCGGCGCTCCGGTCCGGCACGGTGAAGTCG 1468  
QY 221 AlaileGlyLeuGlyLeuLeuSerGluGlyLeuGlyLeuGlyLeuGlyLeuGlyLeu 240  
Db 1467 GCGGTGGGCTGGGCATGCTCTGGCGGAGGAATCGGCACACCATCCGGATTCCCTG 1408  
QY 241 AlaAlaAspProValGluGluLeuLeuValGlyPheAspLeuLeuSerLeuArgile 260  
Db 1407 GCTGCCGATCCCGTCCAGAGATCAAGTTCGATCTCAAGTCCCTGACCTG 1348  
QY 261 ArgSerArgGlyLeuLeuPheileAlaCysProThrCysSerArgGlnGluPheAspVal 280  
Db 1347 CGCTCCCGTGGCATCAACTTCATCGCTGCCCGAGCTGTTCCGGCGAGAACTTCGACGTG 1288  
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspLeuLeuLeuThrProMetAspVal 300  
Db 1287 GTCAAGACCATCAACGAGCTGGAAGGCGCGCTGGAGGACCTGCTGTGTCGATGGACGTG 1228  
QY 301 SerileileGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 1227 GCGGTGATCGGTTCGCTCAACGCTCCGGCGGAAGCAAGAGGCCCATGTGCGCCCTC 1168  
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrgluAspGlyValArgLysAspArgLeuAsp 340  
Db 1167 ACCGGCGGCACCTCCGAACCTG---GTGTATATCGACGGCAAGCGCTCGCAGAACTGACC 1111  
QY 341 AsnAspMetIleAspGlnLeuGluAlaArgileArgAlaLysAlaSerGlnLeuAsp 360  
Db 1110 AACGACAACTGTGTGACGAGCTGGAACGCGTATCCGCGAGAAAGCGCGCAGAGAGGCC 1051  
QY 361 GluAla 362  
Db 1050 GAGGCC 1045

## RESULT 11

US-09-252-991A-6339  
; Sequence 6339, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6339  
; LENGTH: 2088  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6339

## Alignment Scores:

Pred. No.:	2,47e-147	Length:	2088
Score:	1363.50	Matches:	266
Percent Similarity:	86.59%	Conservative:	44
Best Local Similarity:	74.30%	Mismatches:	47
Query Match:	73.07%	Indels:	1
DB:	4	Gaps:	1

US-09-921-992-50 (1-372) x US-09-252-991A-6339 (1-2088)

QY 5 AlaProIleGlnArgArgLysSerThrArgIleTyrgluValGlyAsnValProIleGlyAsp 24  
Db 5 TCTCCGATCATTCGCCGCAAGTCTCGGAAATCTGGGTCCGCAACGTCCTCCGGTGGCGGC 64

QY 25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThr 44  
Db 65 CAGCGCCGATCGCGTGCAGAGCATGACCAACACCGAGACCTGCGACCTGCTGCCACC 124  
QY 45 ValAsnGlnLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64  
Db 125 GTCCGCGAGATCCCGCGCTGGAAGATCCCGCGCGGACATCGTCCGGGTTTCCGTCCC 184  
QY 65 ThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuVal 84  
Db 185 GACATGGACGCGCGCGAGCATTCGGCAAGATCAAGCAGCAGGTCAACGTCGCGTGGTC 244  
QY 85 AlaAspIleHisPheAspTyrglyAlaLeuLysValAlaGluTyrglyValAspCys 104  
Db 245 GCCGACATCCACTTCGACTATCGCATCCCTCGCGCTCGCGAGCTGGAGTGGACTGC 304  
QY 105 LeuArgileAsnProGlyAsnIleGlyAsnGluArgileArgMetValValAspCys 124  
Db 305 CTGCGCATCAATCCGGGCAACATCGGTCCGAGGACCGGGTCAAGCGCGTGTTCATGCC 364  
QY 125 AlaArgAspLysAsnIleProIleArgileGlyValAsnAlaGlySerLeuGluLysAsp 144  
Db 365 GCGCGCGAGCGCAACATCCCGATCGTATCGCGCTCAATCGCGTTCGCTGGAAGGAC 424  
QY 145 LeuGlnGluLysTyrglyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHis 164  
Db 425 CTGCAGAGAAATACGGGCAACCGCACCCCGAAGCCCTGCTCGAATCGGCCATCGGCCAC 484  
QY 165 ValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAsp 184  
Db 485 GTCGATCATCTCGCAAGCTGGACTTCCAGAACTTCAAGGTCAAGGTCAAGGCTCCGAC 544  
QY 185 ValPheLeuAlaValGluSerTyrglyLeuLeuAlaLysGlnIleAspGlnProLeuHis 204  
Db 545 GTCTTCATGGCGCTCGCGCTATCGCTGTGTCGCGCAGCAGCATCGAGCAGCCCTGCAC 604  
QY 205 LeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224  
Db 605 CTGGGCATCACCGAGCGCGCGCTCGCTCCGCAAGTTCGCGCGTGGGGCTG 664  
QY 225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244  
Db 665 GGCATGCTCTCGCGAGGAAATCGCGCACCATCCCGATTTCCCTGCTGCCGATCCG 724  
QY 245 ValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgileArgSerArgGly 264  
Db 725 GTCCGAGAGATCAAGTTCGCTTCGACATCTCCTCAAGTCCCTGCACCTCCCGTGGC 784  
QY 265 IleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrVal 284  
Db 785 ATCAACTTCATCCCTCGCGAGCTGTTTCGCGCGAGAACTTCGACGTGTGTAAGACCATG 844  
QY 285 AsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGly 304  
Db 845 AACGAGCTGGAAGGCGCGCTGGAGGACCTGCTGTGTCGCGATGGACGTGCGCGTATCGGT 904  
QY 305 CysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsn 324  
Db 905 TCGCTGCTCAACGCTCCGGCGCAAGCAAGGAGGCCCATGTCGGCTCCACCGCGCACT 964  
QY 325 LysLysSerGlyLeuTyrgluAspGlyValArgLysAspArgLeuAspAsnAspMet 344  
Db 965 CCGAACCTG---GTGTATATCGACGCAAGCGCTCGCAGAAACTCCACCAACGACCTG 1021  
QY 345 IleAspGlnLeuAlaArgileArgAlaLysAlaSerGlnLeuAspGluAla 362  
Db 1022 GTGGACGAGCTGGAACGCTGATTCGCCAGAGGCGCGCGAGAGCCCGAGGCC 1075

## RESULT 12

US-08-827-190-10/c  
; Sequence 10, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:

```
Db 535 ATTATAGCGCTTCTCCGATCATTCGCGCAAGTCTCGGAAATCTGGGTGCGCAACGTC 594
Qy 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 595 CCGGTGGGGCGGACCGCGCGATCGCGGTGCGAGATGACCAACACCGAGACCTCGAC 654
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 655 GTCCGTGCCACCTCGCGAGATCCCGCGCTGGAAGATCGCGCGCGGACATCTGTCGG 714
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80
Db 715 GTTTCGCTCCCGACATGACGCGCGCGGATTCGCAAGATCAAGCAGCAGGTCAAC 774
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 775 GTGCCGCTGGTCCGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCCCGAGCTG 834
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120
Db 835 GGAGTGGATGCTGCGCATCATTCGGGCAACATCGTCCGAGGACCGGTCAAGGCC 894
Qy 121 ValValAspCysAlaAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 895 GTGGTTCGATGCGCGCGCGGACCAACATCCCGATCCGATCGGTCAATGCCGGTTCG 954
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSer 160
Db 955 CTGGAAAGAGGACTCGCAGAGAAATACGCGCAACCGACCCCGAGCCCTGTCGAATCG 1014
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 1015 GCATGCGCCAGCTCGATCATCTCGCAAGCTGGACTTCAGAACTTCAAGGTCAAGCGTC 1074
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 1075 AAGGCTCCGAGCTCTTCATGCGCGTCCGCGCTATCGCTGCGCAGGACGATCGAG 1134
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
Db 1135 CAGCCCTCTCAGCTGGGCATCACCGAGGCGCGCGCTCGCTCGCGCACGCTGAAGTCG 1194
Qy 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 1195 GCGGTGGGCTGGGATGCTCTGCGCGAGGAATCGCGACACCATCCGGATTTCCCTG 1254
Qy 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 1255 GCTCCGATCCGTCGAGGAGATCAAGGTTCGTTTCGACATCCTCAAGTCCCTGCACCTG 1314
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
Db 1315 CGCTCCGTCGATCAACTTCACTTCATCGCTGCCGAGCTGTTCCGCGCAGAACTTCACG 1374
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 1375 GTGAAGACCATGACGCTGAAGGCGCCCTGGAGGACCTGCTGGTCCGATGACCTG 1434
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 1435 GCGGTGATCGTTCGTCATCAACGTCGCGCGAGCAACCAAGAGGCCCATCTCGGCCTC 1494
Qy 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
Db 1495 ACCGGCGGCACCTCCGAACCTG---GTGTATATCGACGCGCAAGCCGTCGCAGAACTGACC 1551
Qy 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
Db 1552 AACGACACCTGCTGGACGAGCTGGAACCGGTGATCCCGCCAGAGGCGCGCGAGAGGCC 1611
Qy 361 GluAla 362
|||||
```

```
Db 1612 GAGGCC 1617
RESULT 10
US-09-252-991A-6131/c
; Sequence 6131, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6131
; LENGTH: 2202
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6131
Alignment Scores:
Pred. No.: 2,46e-148 Length: 2202
Score: 1372.50 Matches: 267
Percent Similarity: 86.46% Conservative: 46
Best Local Similarity: 73.76% Mismatches: 48
Query Match: 73.55% Indels: 1
DB: 4 Gaps: 1
US-09-921-992-50 (1-372) x US-09-252-991A-6131 (1-2202)
Qy 1 MethAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 2127 ATTATAGCGCTTCTCCGATCATTCGCGCAAGTCTCGGAAATCTGGGTGCGCAACGTC 2068
Qy 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 2067 CCGGTGGCGCGGACGCGCGCATCGCTGCGAGACATGCCAACACCGAGACCTGCGAC 2008
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 2007 GTCCGTGCGCACCGTCGCGACATCCGCCCTCGGAAGATGCCGCGCGGACATCGTCGG 1948
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80
Db 1947 GTTTCGCTCCCGACATGAGCGCGCGGAGCATTCGCAAGATCAACGACGAGTCAAC 1888
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 1887 GTCCGCTGTCGCGGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCCCGAGCTG 1828
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
Db 1827 GGAGTGGACTGCTCCGCTCGCATCAATCCGGGCAACATCGTCCGAGGACCGGGTCAAGGCC 1768
Qy 121 ValValAspCysAlaAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 1767 GTGGTCGATGCGCGCGCGGACCAACATCCCGATCCGATCCGCTCAATGCCGTTTCG 1708
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
Db 1707 CTGGAAAGAGGACTCGCAGAGAAATACGCGCAACCGACCCCGAGACCTGCTCGAATCG 1648
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 1647 GCCATGCGCCACGTCGATCATCTCGCAAGCTGGACTTCCAGAACTTCAAGGTCAAGCGTC 1588
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 1587 AAGGCTCCGAGCTTTCATCGCGCTCGCCCGCTATCGCCCTGCTGGCCAGGACGATCGAG 1528
```



STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/09/170,187  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-09-170-187-7

Alignment Scores:  
Pred. No.: 2,76e-174 Length: 1079  
Score: 1593.50 Matches: 317  
Percent Similarity: 94.93% Conservative: 20  
Best Local Similarity: 89.30% Mismatches: 17  
Query Match: 85.40% Indels: 1  
Gaps: 1

US-09-921-992-50 (1-372) x US-09-170-187-7 (1-1079)

QY 7 IleGlnArgArgLysSerThrArgGlyValGlyAsnValProIleGlyAspGlyAla 26  
DB 6 ATTAAGCGTCTGAATCGACAAATTTATGTGGAAATGTACCAATTTGGTGGGATGCG 65  
QY 27 ProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsn 46  
DB 66 CCTATTGCGTGAATCAATGACAAATCTCGCACCCTGATGTGGAGCGACAGTTGCT 125  
QY 47 GlnIleAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMet 66  
DB 126 CAAATTAATCAATTAGAACGTGTGGTGCAGATATTGTCGTATCTGTTCACCAATG 185  
QY 67 AspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAsp 86  
DB 186 GATGCTCGGAGCAATTAACAAATTAACACAACTGAATGTTCGCTCGTAGCAGAT 245  
QY 87 IleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArg 106  
DB 246 ATTCATTTCGACTATCGTATCGGCTTAAAGTCGACAGATATGGAGTGGATTGTTCAGT 305  
QY 107 IleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArg 126  
DB 306 ATCAATCCTGGCAACATTTGCTGTGAAGATCGCGTCCGCTGTTGTATGTGGCGCA 365  
QY 127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146  
DB 366 GACAAAATATTCCGATTCGTATGTGTAAATGCAGGCTCTTTAGAAAAGATTGCAA 425  
QY 147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166  
DB 426 GAAAAATATGCGCAACCAACCGCAGAACCTTGTAGAAATCCGATTCGCTCATGTAGAA 485

QY 167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186  
DB 486 ATTTAGATCGTCTTAACCTTCGATCAGTTTAAAGTGAGCGTAAAGACCTCCCGATGTTTC 545  
QY 187 LeuAlaValGluSerTyrArgIleLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206  
DB 546 TTAGCGGTTGAATCTATCTGTTTACTGGCTAAGCAATTAACACGCCCTTTCATTTAGCG 605  
QY 207 IleThrGluAlaGlyGlyValAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
DB 606 ATTACAGAAGCAGGTGGCGCAGCGCTGGTGAGTAAATCTGCAGTGGGTTTGAAGAAG 665  
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246  
DB 666 TTATTAGCTGAGGCGCATTTGGCGATACACTACGCGTCTCTTTGGCGGACAGATCCTGTAGAG 725  
QY 247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266  
DB 726 GAAATCAAAGTCGGTTTGTATATTTGAAATCTTTACGGAATTCGTTCAAGAGAATTAAC 785  
QY 267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286  
DB 786 TTTATTTGCTTGCCCAACCTGTTCTGCCAAGAAATTTGATGTAATCGGTACAGTAATGCG 845  
QY 287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306  
DB 846 CTAGAACAACGCTTGAAGATATTATTACCAATGGATGTATCTATTATCGTTGTGTA 905  
QY 307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyAsnLysLys 326  
DB 906 GTGAATGGTCTTGGCGAGGCACTCGTCTCCGATCTCGGCGTAACGGCGGTAAACAAAAA 965  
QY 327 SerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsnAspMetIle 345  
DB 966 ACGGTTATTTATCTTGACGAGAGAACCCCAAAAGAGCGTTTGTATACGAGATATAGTG 1025  
QY 346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
DB 1026 AACCAATTAGAAAGCAAAATTCGTGCGAAAGTCGACGACCAAGAT 1070

## RESULT 9

US-09-252-991A-6420

; Sequence 6420, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6420

; LENGTH: 1644

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6420

Alignment Scores:  
Pred. No.: 1,58e-148 Length: 1644  
Score: 1372.50 Matches: 267  
Percent Similarity: 86.46% Conservative: 46  
Best Local Similarity: 73.76% Mismatches: 48  
Query Match: 73.55% Indels: 1  
Gaps: 1

US-09-921-992-50 (1-372) x US-09-252-991A-6420 (1-1644)

QY 1 MetHisAsnGlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnVal 20

Db 391932 AACAAAAAGCGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTTTGATAACGAA 391991  
QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362  
Db 391992 GATATAGTGACCAATAGNAGCAAAATTCGTGCGAAGTCGCACGACAGATCCAAA 392051  
QY 363 ArgArgIle 365  
Db 392052 AACAGAAATT 392060

## RESULT 7

US-08-827-190-7  
; Sequence 7, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,190  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1079 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-827-190-7

## Alignment Scores:

Pred. No.: 2,76e-174 Length: 1079  
Score: 1593.50 Matches: 317  
Percent Similarity: 94.93% Conservative: 20  
Best Local Similarity: 89.30% Mismatches: 17  
Query Match: 85.40% Indels: 1  
DB: 2 Gaps: 1

US-09-921-992-50 (1-372) x US-08-827-190-7 (1-1079)

QY 7 IleGlnArgLysSerThrArgIleTyrValGlyAsnValProfileGlyAspGlyAla 26  
Db 6 ATTAGCGTCGTAATCGACAAAATTTATGTGGAAATGTACCAATTTGGTGGGATGCG 65  
QY 27 ProfileAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsn 46  
Db 66 CCTATTGCGTCGAATCATGACAAATCTCGACCACTGATGTGGAACGACAGTTGCT 125  
QY 47 GlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMet 66  
Db 126 CAAATTAATATCATTAGAACGTGTGTGCGAGATATTGTTCTGTATCTCTGCCAACATG 185

QY 67 AspAlaAlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAsp 86  
Db 186 GATGCTGCGGAAGCATTTTAAACAAATTTAAACAAAGTCAATGTTCCGCTCGTAGCAGAT 245  
QY 87 IleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArg 106  
Db 246 ATTCATTTTCGACTATCGTATCGGTTAAAGTCGCAGAAATATGGAGTGGATTGTTTACGT 305  
QY 107 IleAsnProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArg 136  
Db 306 ATCAATCTCGCAACATTTGTCGTAAGATCGCGTCGCGCTGCTGTTGATTGTGGCGGA 365  
QY 127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146  
Db 366 GACAAAATATTTCGATTTCGTATTGCTGTAATGCGAGGCTCTTTAGAAAAAGATTGGCAA 425  
QY 147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166  
Db 426 GAAAAATATGCGCAACCAACGCGCAGAAAGCTTTGTTAGAAATCGGCATTGGCTCATGTAGAA 485  
QY 167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValIysAlaSerAspValPhe 186  
Db 486 ATTCTAGATCGTCTTAACCTTCGATCAGTTTAAAGTGAGCGTAAAGCCCTCCGATGATTTC 545  
QY 187 LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206  
Db 546 TTAGCGGTTGAATCTTATGTTTACTGCTAAAGCAATTTAAACAGCCTTTACATTAGGC 605  
QY 207 IleThrGluAlaGlyClyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
Db 606 ATTACAGAAAGAGGTGGCGCACGCTGTGTGAGTAAATCTGCAGTGGGTTTAGGAATG 665  
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaSerProValGlu 246  
Db 666 TTATTAGCTGAGGCAATTTGGCGATACACTACGCTCTCTTTGCGCGCAGATCCTGTAGAG 725  
QY 247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266  
Db 726 GAAATCAAAAGTCGTTTGTATATTTTGAATCTTTACGGATTCGTTCAAGAGGAATTAAC 785  
QY 267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286  
Db 786 TTATTGCTTGCCCAACCTGTTCTGCCAAGAAATTTGATGTAATCGGTACAGTAAATGCG 845  
QY 287 LeuGluGlnArgLeuGluAspIleThrProMetAspValSerIleIleGlyCysVal 306  
Db 846 CTAGAACCAACGCTTTGAAGATATTATACCAATGGATGTATCTATTATTCGTTGTGTA 905  
QY 307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLys 326  
Db 906 GTGAATGTCCTGGCGAGGCACCTCTCTCCGATCTCGCGGTAAACGGCGGTAAACAAAAA 965  
QY 327 SerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAspMetIle 345  
Db 966 AGCGGTATTATCTTACGCGAGAACGCCAAAAGAGCGTTTTTGATAACGAAGATATAGTG 1025  
QY 346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
Db 1026 AACCAATTAGAGCAAAATTCGTGCGAAAGTCGCACGACCAAGAT 1070

## RESULT 8

US-09-170-187-7  
; Sequence 7, Application US/09170187  
; Patent No. 6383745  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco

Db 391812 GTAAATGCGCTAGAACACCGCTTGAAGATATTATTACCAATGATGATCTATTATC 391871  
 QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323  
 Db 391872 GGTGTGTGTGTAATGCTGCGGAGGCACTCGTCTCCGATCTCGCGTAACGGCGGT 391931

QY 324 AsnLysLysSerGlyLeuValGluAspGlyValArg--LysAspArgLeuAspAsn 342  
 Db 391932 AACAAAAAGCGGTATTATCTTGACGGAGACGCCAAGAGCGTTTGTATACGAA 391991  
 QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362  
 Db 391992 GATATAGTGAACCAATTAGAACCAAAAATTCGTGCGAAAGTCCGACGACAAAGTCCAAAA 392051

QY 363 ArgArgIle 365  
 Db 392052 AACAGAAATT 392060

RESULT 6  
 US-09-643-990A-1  
 ; Sequence 1, Application US/09643990A  
 ; Patent No. 6528289  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Robert D. Fleischmann  
 ; Mark D. Adams  
 ; Owen White  
 ; Hamilton O. Smith  
 ; J. Craig Venter  
 ; TITLE OF INVENTION: The Nucleotide sequence of  
 ; the Haemophilus influenzae Rd Genome, Fragments  
 ; Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Human Genome Sciences, Inc.  
 STREET: 9410 Key West Avenue  
 CITY: Rockville,  
 STATE: MD  
 COUNTRY: USA  
 ZIP: 20850  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: 3 1/2 inch diskette  
 COMPUTER: Dell Pentium  
 OPERATING SYSTEM: MS DOS v6.22  
 SOFTWARE: ASCII Text  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/643,990A  
 FILING DATE: 23-Aug-2000  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/487,429  
 FILING DATE: 1995-06-07  
 APPLICATION NUMBER: 08/426,787  
 FILING DATE: 1995-04-21  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kenley K. Hoover  
 REGISTRATION NUMBER: 40,302  
 REFERENCE/DOCKET NUMBER: PB186P1C1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 301-610-5790  
 TELEFAX: 310-309-8439  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1830121 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1  
 Alignment Scores:  
 Pred. No.: 1.38e-170 Length: 1830121  
 Score: 1603.50 Matches: 320

Percent Similarity: 93.66% Conservative: 20  
 Best Local Similarity: 88.15% Mismatches: 22  
 Query Match: 85.93% Indels: 1  
 DB: 4 Gaps: 1  
 US-09-921-992-50 (1-372) x US-09-643-990A-1 (1-1830121)  
 QY 4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23  
 Db 390972 CAGCCAACCTATTAAAGCTCGTAATCGACAAAATTTATGTGGAAATGTACCAATTGGT 391031  
 QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43  
 Db 391032 GGGGATGCGCTATTGCGGTGCAATCAATGACAAATACTCGCACCACTCATGTGAAGCG 391091  
 QY 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63  
 Db 391092 ACAGTTGCTCAAAATTAATCATTAGAACGTGTGTGGCAGATATTGTCGTGTATCTGTT 391151  
 QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83  
 Db 391152 CCACAAATGGATGCTGCGGAGCATTTAAACAAATTAACAAAGTGAATGTTCCGCTC 391211  
 QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103  
 Db 391212 GTAGCAGATATTCAITTCGACTATCGTATCGGTTAAAAAGTCCGAGAATATGGAGTGGAT 391271  
 QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123  
 Db 391272 TGTTCAGCTATCAATCTCGCAACATTTGCTGGAAGATCGCGTCCGCTGCTGTTGTGAT 391331  
 QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySerLeuGluLys 143  
 Db 391332 TGTGCGGAGACAAAAATATTCGATTCTGATTGGTGAATGCAGGCTCTTTAGAAAAA 391391  
 QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163  
 Db 391392 GATTTGCAAGAAAAATATGCGAACCCACGCGCAGAACGCTTGTAGAAATCCGATTCGCT 391451  
 QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183  
 Db 391452 CATGTAGAAAATCTTAGATCTCTTAACCTCGATAGTAAAGTGAAGCGTAAAGGCTCC 391511  
 QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203  
 Db 391512 GATGATTTCTTAGCGGTGAATCTTATCGTTTACTGGCTAAACCAATTAAACAGCCTTTA 391571  
 QY 204 HisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223  
 Db 391572 CATTAGGCATTACAGAACGAGTGGCGCACGCGCTGGTGCAGTAAATCTGCAGTGGGT 391631  
 QY 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243  
 Db 391632 TTAGGAATGTTATTAGCTAGGGCATTTGCGGATACACTACGCGTCTCTTTGGCGGAGAT 391691  
 QY 244 ProValGluGluLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263  
 Db 391692 CCTGTAGAGAAATCAAGTCGTTTGTATATTTTGAATCTTTACGGATTCGTTCAAGA 391751  
 QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGlnPheAspValIleGlyThr 283  
 Db 391752 GGAATTTAACTTTATGCTTGTGCCAACCTGTTCTCGCCAAAGATTTGATGTAATCGGTACA 391811  
 QY 284 ValAsnAlaLeuGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303  
 Db 391812 GTAAATGCGCTAGAACCAACGCTTGAAGATATTATTACCAATGATGATCTATTATTC 391871  
 QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323  
 Db 391872 GGTGTGTGTGTAATGCTGCGGAGGCACTCTCTCCGATCTCGCGTAACGGCGGT 391931  
 QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsn 342

Db 775 CAACCACTTCACCTCGGTATTACAGAACGGGTGGGCTCGTTCTGTTTCAGTCAAAATCA 834  
QY 221 AlaileGlyLeuGlyLeuLeuSerGluGlyLeuGlyLeuGlyLeuGlyLeuGlyLeuGlyLeu 240  
Db 835 GCAATTGGTCTTGTTGGTGAAGGATCGCGCATACGTTACGTATCTCACTC 894  
QY 241 AlaAlaAspProValGluGluLeuLysValGlyPheAspLeuLysSerLeuArgile 260  
Db 895 GCGGAGATCCTGTTGAGGAAGTGAAGTGGTTTGGATTTCTAAATCTTAAATCTTACGGATC 954  
QY 261 ArgSerArgGlyLeuLeuPheLeuAlaCysProThrCysSerArgGlnGluPheAspVal 280  
Db 955 CGCTACGTGGCATCAACTTATTGCTTGCACCACTGTTTCACGCCCAAGAATTGATGTG 1014  
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspLeuLeuThrProMetAspVal 300  
Db 1015 ATTGGTACGGTAATGCTTTGGAGCAGCGCTCGAAGATATTATCACGCCGATGGATGC 1074  
QY 301 SerIleileGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 1075 TCTATTATTGTTGTAGTGAATGCGCGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134  
QY 321 ThrGlyCysAsnLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeu 339  
Db 1135 GCTGGCGCCGAAACCAAAAGTGGTTCTATGAAGATGGCGTTCGCAAAAAGAGCGTTT 1194  
QY 340 AspAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeu 359  
Db 1195 GATAATGACAAATATTATGATCAGCTTGAGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254  
QY 360 AspGluAlaArgArgileAspValGlnGlnValGlu 371  
Db 1255 GATGAAT-AACCGTATAAGATAACCAAGTGCAG 1289

RESULT 5

US-09-557-884-1  
; Sequence 1, Application US/09557884  
; Patent No. 6506581  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: The Nucleotide sequence of  
; the Haemophilus influenzae Rd Genome, Fragments  
; Thereof, and Uses Thereof  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: MD  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3 1/2 inch diskette  
; COMPUTER: Dell Pentium  
; OPERATING SYSTEM: MS DOS v6.22  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/557,884  
; FILING DATE: 25-Apr-2000  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/476,102  
; FILING DATE: JUN-5-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michelle S. Marks  
; REGISTRATION NUMBER: 41,971  
; REFERENCE/DOCKET NUMBER: PB186P3  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 301-309-8504  
; TELEFAX: 301-309-8439  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1830121 base pairs

; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-557-884-1  
Alignment Scores:  
Pred. No.: 1,38e-170 Length: 1830121  
Score: 1603.50 Matches: 320  
Percent Similarity: 93.66% Conservatives: 20  
Best Local Similarity: 88.15% Mismatches: 22  
Query Match: 85.93% Indels: 1  
DB: 4 Gaps: 1  
US-09-921-992-50 (1-372) x US-09-557-884-1 (1-1830121)  
QY 4 GlnAlaProIleGlnArgGlySerThrArgIleTyrValGlyAsnValProIleGly 23  
Db 390972 CAGCCAACTATTAAAGCGTCTGAATCGACAAAATTTATGTGGAAATGTACCAATTTGGT 391031  
QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43  
Db 391032 GGGGATGCCCTATTTCCTGCAATCAATGACAAATACTCGCACCACTGATGTGGAAGCG 391091  
QY 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63  
Db 391092 ACAGTTGCTCAATTAATCATTTAGAACGTTGTTGGTCAGATATTGTTCTGTATCTGTT 391151  
QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83  
Db 391152 CCAACAATGGATGCTGCGGAAGCATTTAAACAAATTTAAACAAACAACTGAATGTTCCGCTC 391211  
QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103  
Db 391212 GTAGCAGATATTTCATTTTCGACTATCGTATCGGTTTAAAGTGCAGAAATATGGAGTGGAT 391271  
QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123  
Db 391272 TGTTTACGTATCAATCTCGCAACATTTGCTGGAAGATCGCGTCCGTGCTGCTGTTGTTGAT 391331  
QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143  
Db 391332 TGTGCGCGAGACAAAATATTTCGATTCGTTATTTGGTAAATGCAGGCTCTTTAGAAAAA 391391  
QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163  
Db 391392 GATTTGCAAGAAAAATATGCGCAACCAACCGCAGAACGCTTCTTAGAAATCCGATTCGCT 391451  
QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183  
Db 391452 CATGTAGAAATTTCTAGATCGTCTTAACATCGATCTTAAAGTGAAGCCTAAAGCCTCC 391511  
QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203  
Db 391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTACTGGCTAAAGCAATTAACAGCCTTTA 391571  
QY 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223  
Db 391572 CATTTAGGCATTACAGAAGCAGGTGGCGCACGCGTGGTGCAGTAAATCTCGAGTGGGT 391631  
QY 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243  
Db 391632 TTAGGAATGTTATTAGCTGAGGGCAATTCGCGATACACACGCGCTCTCTTTGCGGCGAGAT 391691  
QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263  
Db 391692 CTTGTAGAGAAATCAAAAGTCGTTTGTATATTTTGAATCTTTACGGATTCGTTCAAGA 391751  
QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283  
Db 391752 GGAATTAACCTTTATTGCTTGCCCAACCTGTTCTCGCAAGAAATTTGATGTAATCGGTACA 391811  
QY 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303

161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheValSerVal 180  
655 GCATCGCATGTTGATATCTTGGACAGCTGAATTCATCAGTCAAGGTCAAGTGT 714  
181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
715 AAAGCGTCGATGCTCTTCTGCGCTGCTTATCGTTTATTTGGCGCAAAATTTGAT 774  
201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220  
775 CAACCACTTCACCTCGGTATTCAGAAAGCGGTGGGCTCGTTCTGTTCAAGTAAATCA 834  
221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
835 GCAATTTGGTCTTGGTATGTTGGCTGAAGGTATCGCGATACGTTACGTATCTCCTC 894  
241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
895 GCGGCAGATCTCTGAGGAAGTAAAGTCGGTTTGTATTTCTAAATCGTTACGGATC 954  
261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
955 CGCTCAGTGGCATCAACTTATTTGCTTGCCCAACCTGTTCAAGCAAGAAATTTGATG 1014  
281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
1015 ATTGGTACGGTAAATGCTTTGGAGCAGCGCTCGAAGATATTATCAGCGCGATGATG 1074  
301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
1075 TCTATTATTGTTGTTAGTATGATGATGCGCGGTGAAGCGAGGTTCTACTTTAGTGTG 1134  
321 ThrGlyCysAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeu 339  
1135 GCTGGCGCGCAAAACCAAAAGTGGTTCTATGAAGATGGCGTTCGCAAAAGAGCGTTT 1194  
340 AspAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeu 359  
1195 GATAATGCAATATATTATGATCAGTTGAGCGCAAAATTCGCGCAAAAGCGCAATGCTT 1254  
360 AspGluAlaArgIleAspValGlnGlnValGlu 371  
1255 GATGAAT-AACGTTAATAGATAACCAAGTCGAA 1289

## RESULT 4

US-09-170-187-4  
; Sequence 4, Application US/09170187  
; Patent No. 6383745  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlin & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/170,187  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/827,190  
; FILING DATE:

## ATTORNEY/AGENT INFORMATION:

; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443

## TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:

## SEQUENCE CHARACTERISTICS:

; LENGTH: 1345 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:

## NAME/KEY: CDS

; LOCATION: 175..1272

; OTHER INFORMATION: /gene= "aarC"

; OTHER INFORMATION: /note= "Similar to E. coli GcpE protein listed by GenBank."  
US-09-170-187-4

## Alignment Scores:

Pred. No.: 1,02e-177 Length: 1345  
Score: 1624.50 Matches: 325  
Percent Similarity: 94.09% Conservative: 25  
Best Local Similarity: 87.37% Mismatches: 21  
Query Match: 87.06% Indels: 2  
DB: 4 Gaps: 1

US-09-921-992-50 (1-372) x US-09-170-187-4 (1-1345)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
Db 175 ATGCATATGAATCACCAGATAAAGACGTAATAATCCACCCGAATTTATGAGTAAACG 234  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
Db 235 CCTATTGGCGATGGTCTCCCATTTGCTGTCCTATATGACGAATACGCGACGCGAT 294  
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
Db 295 GTTGAAGCCACCTGTGCGGCAAAATCCCAATCATTGACGCTGTAGGTGTTGATATCGTCC 354  
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsn 80  
Db 355 GTGTCGTCTTCTACGATGATGATGACGAGAAAGCCTTTAAATTTAATTAAAGCAGCGTGAAT 414  
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
Db 415 GTGCCATTGTTGGCGATATTCACTTTGACTACCGTATCGCGATGAAAGTGGCTGAATAT 474  
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
Db 475 GGTGTTGACTGCTACGAATTAACCCAGGTAATATATCGGAGTGAAGAGCGTATTTCGCCAA 534  
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
Db 535 GTGCTTGATGATGCTCGTCATCACACATTCCTATCCGATATAGGGTCAATGCGCGGTCA 594  
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
Db 595 CTGGAAAAAGATATCAAGAAAAATACGCTGAGCAACACCTGAACGATTGTTGAATCA 654  
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
Db 655 GCAATCGACATGTTGATATCTTGGACAGGCTGAATTTCCAGTTCAGTTCAGGTGTT 714  
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
Db 715 AAAGCGTCGATGCTCTTCTGCGCTGCTTATCGTTTATTTGGCGCAAAATTTGAT 774  
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220

Db 121 GTGGAAGCAACGGTCAATCAATCAAGCGCTGCGCTGCGCTGATATCGTCCGT 180  
Qy 61 ValSerValProThrMetAspAlaGluAlaPheLysLeuLysGlnValAsn 80  
Db 181 GTATCCGTACCGAGATGGACGGGAGAGCGCTCAAACTCATCAACAGAGGTAAAC 240  
Qy 81 ValProLeuValAlaAspLysHisPheAspTyrArgLysValAlaGluTyr 100  
Db 241 GTGCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTAGGCGAATAC 300  
Qy 101 GlyValAspCysLeuArgLysAsnProGlyAsnLysGluArgLysMet 120  
Db 301 GCGCTCGATTTGCTGGTATTAACCTCGGCAATATCGTAATGAAGAGCGTATTCGCATG 360  
Qy 121 ValValAspCysAlaArgLysAsnLysProLysArgLysValAlaGluTyr 140  
Db 361 GTGGTTGACTGTGCGCGGCAATAAAACATTCGATCCGATTCGCTTAAACGCGGATCG 420  
Qy 141 LeuGluLysAspLeuGlnLysTyrGlyLysProThrProGlnAlaLeuLeuGluSer 160  
Db 421 CTGGAAAAGATCTGCAAGAAAGTATGGCAACCGACCGCGGCTGCTGGATCT 480  
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
Db 481 GCATCGCTCATGTTGATCATCTCGATCGCTGCACTCGATTCGATTCAGTTCAAAGTCAGCGTG 540  
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnLysAsp 200  
Db 541 AAAGCTCTGACGCTCTTCCTCGCTGTGAGTCTTATCGTTGTCGCAAAACAGATCGAT 600  
Qy 201 GlnProLeuHisLeuGlyLysThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220  
Db 601 CAGCCGTTGATCTGGGGATCACCGAAGCGGTTGGCGGACGGCGGCGATTAATCC 660  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyLysGlyAspThrLeuArgValSerLeu 240  
Db 661 GCATTTGGTTAGTCTGCTGCTGTCTGAAGGATCGCGGACATCGCGCTATCGCTG 720  
Qy 241 AlaAlaAspProValGluLysLysValGlyPheAspLysLeuLysSerLeuArgLys 260  
Db 721 GCGGCGATCCGCTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCGCTGCGTATC 780  
Qy 261 ArgSerArgGlyLysAsnPheLysAlaCysProThrCysSerArgGlnGluPheAspVal 280  
Db 781 CGTTCGAGGAGATCACTTCATCGCTGCGGACCTGTTCGGCTCAGGAATTTGATGT 840  
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspLysLeuThrProMetAspVal 300  
Db 841 ATCGGTACGGTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGGACGT 900  
Qy 301 SerIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 901 TCGATTTATCGCTGCTGTGTAATGCCCGGAGGCGTGGTGTCTACACTCGCGCTC 960  
Qy 321 ThrGlyLysLysLysSerGlyLeuTyrGluAspGlyValArgLys 336  
Db 961 ACCGGCGCAACAAAGAAAGCGCCCTCTATGAAGATGGCGTGGCAAA 1008

## RESULT 3

US-08-827-190-4

; Sequence 4, Application US/08827190

; Patent No. 5858367

; GENERAL INFORMATION:

; APPLICANT: Rather, Philip N.

; TITLE OF INVENTION: Methods For Screening For Antimicrobials

; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof

; NUMBER OF SEQUENCES: 13

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Medlen &amp; Carroll, LLP

; STREET: 220 Montgomery Street, Suite 2200

; CITY: San Francisco

; STATE: California

; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,190  
; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Carroll, Peter G.

; REGISTRATION NUMBER: 32,837

; REFERENCE/DOCKET NUMBER: CASE-02443

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1345 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 175..1272

; OTHER INFORMATION: /gene= "aarC"

; OTHER INFORMATION: /note= "Similar to E. coli GcpE protein listed by GenBank."  
US-08-827-190-4

## Alignment Scores:

Pred. No.: 1,028-177 Length: 1345  
Score: 1624.50 Matches: 325  
Percent Similarity: 94.09% Conservative: 25  
Best Local Similarity: 87.37% Mismatches: 21  
Query Match: 87.06% Indels: 2  
DB: 2 Gaps: 1

US-09-921-992-50 (1-372) x US-08-827-190-4 (1-1345)

Qy 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
Db 175 ATGCATATGATGATACCGGATATAAAGAGAGCTAATCCACCCGGAATTTATGTAGTAACTG 234  
Qy 21 ProfileGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
Db 235 CCTATTGGCGATGGTGTCTCCCATTCGTCTCAATCTATGACGAATACGCGCAGCGGAT 294  
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
Db 295 GTTGAAGCCACTGTGCGGCAATCCCAATCACTTGAAGCGTGTAGGTGTTGATATCGTCCG 354  
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuLysGlnGlnValAsn 80  
Db 355 GTGTCTGTCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 414  
Qy 81 ValProLeuValAlaAspLysHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
Db 415 GTGCCATTGGTTGGGATATTTCATTTGACTACCGTATCGCGATGAAAGTGGCTGAATAT 474  
Qy 101 GlyValAspCysLeuArgLysAsnProGlyAsnLysGluArgLysMet 120  
Db 475 GGTGTGTGATGCTGCTACGAAATTAACCCAGGTAATATCGGAGTGAAGAGCGTATTCGCCAA 534  
Qy 121 ValValAspCysAlaArgLysAsnLysProIleArgIleGlyValAsnAlaGlySer 140  
Db 535 GTCGTTGATGATGCTGCTGATCATCAACATTCCTATCCGTTATAGGGGTCAATGGCGGTCA 594  
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
Db 595 CTGGAAAAAGATATCCAAAGAAAAATACGGTGAGCCAAACACCTGAAGCATTTGTTGAATCA 654

```
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-827-190-8

Alignment Scores:
Pred. No.: 1,43e-185 Length: 1010
Score: 1691.00 Matches: 336
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.62% Indels: 0
DB: 2 Gaps: 0

US-09-921-992-50 (1-372) x US-08-827-190-8 (1-1010)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATAACAGGCTCCAAATTCACGTAGAAAATCAACACGATTTACGTTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 61 CCATTTGGCGATGCTGCTCCATCGCGGTACGTACGTCCATGACCAATACGGGTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsn 80
DB 181 GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACGAGGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTCCCGCTGGTGGCTGACATCCACTTCGATATCGCATTCGCTGAAAGTAGGCGGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120
DB 301 GCGCTCGATTGCTGGGTATTAACCCCTGGCAATATCGTAATGAAGAGCGTATTCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAlaGlySer 140
DB 361 GTGGTTGACTGTCGCGCGGATAAACATTCGATCCGATTCGCTTATCGCGTTAACGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyLysProThrProGlnAlaLeuLysGluSer 160
DB 421 CTGGAAGAGATCTGCAAGAAAGATATGGCAACCGCGCGGCGGCTTGCTGGATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGTGTGACGCTCTTCCTCGCTGTGTAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 601 CACCGGTTGATCTGGGGATCACCGAAGCCGGTGGTGGCGGCGAGCGGGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGTTTAGTCTGCTGCTGCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGGATCCGTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCGCTGCGGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCCGGAGGATCAACTTCATCGCTGCCACCTGTTCCGCTCAGGAATTTGATGTT 840
QY 281 IleGlyThrValAlaAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
```

```
DB 841 ATCGGTACGTTTAAAGCGCTGGAGCAACGCCCTGGAGATATCATCATCCGATGGACGTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 901 TCGATTATCGCTGCTGCTGATGATGCCCGAGTGAGCGCTGTTTCTACACTCGCGGTC 960
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLys 336
DB 961 ACCGGCGCAACAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAA 1008

RESULT 2
US-09-170-187-8
; Sequence 8, Application US/09170187
; Patent No. 6383745
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-09-170-187-8

Alignment Scores:
Pred. No.: 1,43e-185 Length: 1010
Score: 1691.00 Matches: 336
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.62% Indels: 0
DB: 4 Gaps: 0

US-09-921-992-50 (1-372) x US-09-170-187-8 (1-1010)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATAACAGGCTCCAAATTCACGTAGAAAATCAACACGATTTACGTTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 61 CCATTTGGCGATGCTGCTCCATCGCGGTACGTACGTCCATGACCAATACGGGTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAsn 80
DB 181 GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACGAGGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTCCCGCTGGTGGCTGACATCCACTTCGATATCGCATTCGCTGAAAGTAGGCGGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120
DB 301 GCGCTCGATTGCTGGGTATTAACCCCTGGCAATATCGTAATGAAGAGCGTATTCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAlaGlySer 140
DB 361 GTGGTTGACTGTCGCGCGGATAAACATTCGATCCGATTCGCTTATCGCGTTAACGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyLysProThrProGlnAlaLeuLysGluSer 160
DB 421 CTGGAAGAGATCTGCAAGAAAGATATGGCAACCGCGCGGCGGCTTGCTGGATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGTGTGACGCTCTTCCTCGCTGTGTAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 601 CACCGGTTGATCTGGGGATCACCGAAGCCGGTGGTGGCGGCGAGCGGGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGTTTAGTCTGCTGCTGCTGAAGGCATCGGCGACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGGATCCGTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCGCTGCGGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCCGGAGGATCAACTTCATCGCTGCCACCTGTTCCGCTCAGGAATTTGATGTT 840
QY 281 IleGlyThrValAlaAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 23, 2003, 17:21:14 ; Search time 76 Seconds  
(without alignments)  
2160.454 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAPIQRKSTRIYGVN.....RAKASQLDEARRIDVQVEK 372

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09921992/runat\_21112003\_162547\_18433/app\_query.fasta\_1.519  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPECL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -ENDS=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09921992 -CGN\_1\_56\_orunat\_21112003\_162547\_18433 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1691	90.6	1010	2	US-08-827-190-8
2	1691	90.6	1010	4	Sequence 8, Appli
3	1624.5	87.1	1345	2	US-08-827-190-4
4	1624.5	87.1	1345	4	Sequence 4, Appli
5	1603.5	85.9	1830121	4	US-09-170-187-4
6	1603.5	85.9	1830121	4	US-09-557-884-1
7	1593.5	85.4	1079	2	US-09-643-990A-1
8	1593.5	85.4	1079	4	Sequence 1, Appli
9	1372.5	73.6	1644	4	US-08-827-190-7
10	1372.5	73.6	1644	4	Sequence 7, Appli
11	1363.5	73.1	2088	4	US-09-252-991A-6420
12	1363.5	73.1	2088	4	Sequence 6420, Ap
					Sequence 6131, Ap
					Sequence 6339, Ap
					Sequence 10, Appli

C 13	1353	72.5	886	4	US-09-170-187-10	Sequence 10, Appli
C 14	1195	64.0	1137	4	US-09-328-352-3780	Sequence 3780, Ap
C 15	812	43.5	886	2	US-08-827-190-9	Sequence 9, Appli
C 16	812	43.5	886	4	US-09-170-187-9	Sequence 9, Appli
C 17	764	40.3	555	4	US-09-252-991A-6251	Sequence 6251, Ap
C 18	751.5	40.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 19	751.5	40.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 20	607.5	32.6	543	4	US-09-252-991A-6209	Sequence 6209, Ap
C 21	470.5	25.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
C 22	288.5	15.5	1083	4	US-09-221-017B-309	Sequence 309, App
C 23	112.5	6.0	3177	4	US-09-252-991A-12534	Sequence 12534, A
C 24	111.5	6.0	3998	3	US-09-066-046-5	Sequence 5, Appli
C 25	111	5.9	1665	4	US-09-221-017B-1017	Sequence 1017, Ap
C 26	106.5	5.7	1854	4	US-09-252-991A-14350	Sequence 14350, A
C 27	106.5	5.7	2943	4	US-09-252-991A-14267	Sequence 14267, A
C 28	106	5.7	2259	4	US-09-252-991A-784	Sequence 784, App
C 29	106	5.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 30	105.5	5.7	2129	3	US-08-975-762-39	Sequence 39, Appli
C 31	105.5	5.7	2129	3	US-09-295-028-39	Sequence 39, Appli
C 32	105.5	5.7	2129	4	US-09-106-582-39	Sequence 39, Appli
C 33	103.5	5.5	2214	1	US-07-985-458-1	Sequence 1, Appli
C 34	101.5	5.4	1694	3	US-09-136-421-11	Sequence 11, Appli
C 35	101.5	5.4	2989	1	US-07-927-851-1	Sequence 1, Appli
C 36	101.5	5.4	2989	1	US-08-453-323-1	Sequence 1, Appli
C 37	101.5	5.4	2989	1	US-08-440-520-1	Sequence 1, Appli
C 38	101	5.4	1149	4	US-09-252-991A-13125	Sequence 13125, A
C 39	101	5.4	1398	4	US-09-252-991A-12815	Sequence 12815, A
C 40	101	5.4	1572	4	US-09-612-964-1	Sequence 1, Appli
C 41	100.5	5.4	1863	4	US-09-328-352-3604	Sequence 3604, Ap
C 42	100.5	5.4	2010	3	US-09-039-859-1	Sequence 1, Appli
C 43	100.5	5.4	11601	2	US-08-222-617A-3	Sequence 3, Appli
C 44	100.5	5.4	11601	2	US-08-222-617A-24	Sequence 24, Appli
C 45	99.5	5.3	3414	4	US-09-252-991A-8258	Sequence 8258, Ap

#### ALIGNMENTS

RESULT 1  
US-08-827-190-8  
; Sequence 8, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aaC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,190  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,937  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1010 base pairs  
; TYPE: nucleic acid





## COMMENT

Contact: Ross Whetten  
Forest Biotechnology Group  
North Carolina State University  
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
NC, 27695-8008  
Tel: 919-515-7800  
Fax: 919-515-7801  
Email: rosswhet@unity.ncsu.edu  
Seq primer: 5' lambda Triplex2 Sequencing Primer.

## FEATURES

source  
1..601  
/organism="Pinus taeda"  
/mol\_type="mRNA"  
/db\_xref="taxon:3352"  
/clone="ST23E04"  
/lab\_host="E. coli BM25.8"  
/clone\_lib="Pine Triplex shoot tip library"  
/notes="Organ: shoot tips; Vector: Lambda Triplex; Site1:  
SfiI (A); Site2: SfiI (B); Shoot tips (approx. 2 cm from  
apex) were collected during the spring, frozen and used  
for mRNA isolation. The SMART-PCR method (Clontech) was  
used to prepare a library from 1 ug total RNA, using the  
Lambda Triplex vector. Plasmid subclones in pTriplex were  
recovered by cre-lox excision in E. coli strain BM25.8 and  
sequenced from the 5' end."  
BASE COUNT 162 a 103 c 166 g 165 t 5 others  
ORIGIN

Alignment Scores:  
Pred. No.: 7,76e-20 Length: 601  
Score: 262.00 Matches: 57  
Percent Similarity: 62.30% Conservative: 19  
Best Local Similarity: 46.72% Mismatches: 40  
Query Match: 14.04% Indels: 6  
DB: 9 Gaps: 2

US-09-921-992-50 (1-372) x AW042702 (1-601)

QY 132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrglyGlu 151  
Db 24 ATCCGAATTGGCACAACCATGGAGTCTTCGATCGTACTATAGTATTATATGTGTAT 83  
QY 152 ProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171  
Db 84 ----TCTCCAGGGGTATGGTGAATCAGCATTTGAATTTCCAGCATTTGCCGAGATTTG 140  
QY 172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191  
Db 141 GGTTTTCATAATTTGTGTGTTTTCATGAAGCGAGCGATCCTGTAGTCATGTTTCAGGCA 200  
QY 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206  
Db 201 TACCGTTTACTTGTTCGGAGATGTATGTGCAAGGATGGGATTTATCCATTGCATTAGGA 260  
QY 207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
Db 261 GTTACTGAGCTGTGAGGTGAAGATGACGATGAGTCTGCATTGCGATTGGAACA 320  
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246  
Db 321 CTTTTCAGGATGTTTGGGTGATACTATTTCGAGTTTCCCTTACAGAACCTCCAGAGAG 380  
QY 247 GluIle 248  
Db 381 GAGATC 386

## RESULT 15

BJ551332 603 bp mRNA linear EST 14-NOV-2002  
LOCUS  
DEFINITION  
BJ551332 K. Sato unpublished cDNA library, strain H602 adult,  
heading stage top three leaves Hordeum vulgare subsp. spontaneum  
cDNA clone bah59b20 5', mRNA sequence.  
ACCESSION  
BJ551332

## VERSION

BJ551332.1 GI:24969783

## KEYWORDS

Hordeum vulgare subsp. spontaneum

## SOURCE

Hordeum vulgare subsp. spontaneum  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae  
; Triticeae; Hordeum.

## REFERENCE

1 (bases 1 to 603)

## AUTHORS

Sato, K., Saisho, D. and Takeda, K.

## TITLE

Barley EST sequencing project in NIG and Okayama Univ

## JOURNAL

## COMMENT

Unpublished  
Contact: Tadao Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

## FEATURES

source  
1..603  
/organism="Hordeum vulgare subsp. spontaneum"  
/mol\_type="mRNA"  
/strain="H602"  
/db\_xref="taxon:77009"  
/clone="bah59b20"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, strain H602  
adult, heading stage top three leaves"  
BASE COUNT 156 a 112 c 169 g 165 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 1.02e-19 Length: 603  
Score: 261.00 Matches: 56  
Percent Similarity: 63.93% Conservative: 22  
Best Local Similarity: 45.90% Mismatches: 38  
Query Match: 13.99% Indels: 6  
DB: 12 Gaps: 2

US-09-921-992-50 (1-372) x BJ551332 (1-603)

QY 132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrglyGlu 151  
Db 42 CTGCGTATCGGACAAATCATCGTAGTCTTTCTGACCGGATATAGCTACTATGTGTAT 101  
QY 152 ProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171  
Db 102 --TCTCCAAGGGAATGGTTCAGTCTGCTTTGGAATTTGCTAGGATCTGTCCGAATTTG 158  
QY 172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191  
Db 159 GACTTCCATAACTTTGTATTTTCAATGAAGCAAGTAACCCCTGTTGTCTATGTGCCAAGA 218  
QY 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206  
Db 219 TATCCGCTCTTGTACGGAATGTATTAACCTTGGATGGGATTTATCCTTTGCACITGGGA 278  
QY 207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226  
Db 279 GTTACCGAAGCTGGTTCAGGGTGAAGATGGGAGGATGAGTCTGCTATTGCGCATCGGAACA 338  
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246  
Db 339 CTTTTCAGGATGTTTGGGTGATACAAATCCGTGTATCCCTTACAGAACCCAGAGGAA 398  
QY 247 GluIle 248  
Db 399 GAAAT 404

Search completed: November 23, 2003, 19:28:18  
Job time : 2608 secs

QY 90 AsPTyArgIleAlaLeuLysValAlaGluTyArgGlyValAspCysLeuArgIleAsnPro 109  
 DB 13 GACAGCGGCCAGTTGAGCAATTTAGAGTACACAGAGATGACTATCAGAAGAACTC 72  
 QY 110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129  
 DB 73 GAGCATATT-----GAGGAGGTTTTCACACATTCGTCTCAACCAATGCTGTAAGAGTATGGA 126  
 QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLysTyTyr 149  
 DB 127 CGTGCATATCGCATTTGGGCAACCACTTGGAGGCTTTCAGATCGCATTTAGAGCTATTAT 186  
 QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169  
 DB 187 GGGGAC---TCGCCTAGGGGAATGGTAGATCAGCATTTTCAGTTTCAAGAATTTGTAGA 243  
 QY 170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189  
 DB 244 AAGTTGGACTTTCACAAATTCGTCTCTCAATGAAGCTAGCAATCCAGTAGTTATGGTT 303  
 QY 190 GluSerTyArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHis 204  
 DB 304 CAGCGATATCGCCTCTTGTAGCTGAGATGTATGTTCAAGGATGGGACTATCCTTTACAC 363  
 QY 205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224  
 DB 364 TTGGGAGTACTGAAGCTGTGAGGTGAGGATGGAAGATGCTGCAATCGGTATT 423  
 QY 225 GlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244  
 DB 424 GGCACACTTCTTCAGGATGTTTGGGTGATACAAATACGGGTTTCCCTCACTGAAGCTCCA 483  
 QY 245 ValGluGluLeu 248  
 DB 484 GAAGAGGAGATA 495  
 RESULT 13  
 ID AU186794 standard; RNA; EST; 489 BP.  
 AC AU186794;  
 SV AU186794.1  
 XX 23-MAY-2003 (Rel. 75, Created)  
 DT 23-MAY-2003 (Rel. 75, Last updated, Version 1)  
 XX Porphyra yezoensis cDNA, clone:PF004b08\_r, 5'end.  
 DE EST (expressed sequence tag).  
 KW Porphyra yezoensis  
 XX Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
 [1]  
 RN Asamizu E., Nakamura Y.;  
 RA Submitted (15-JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RL Erika Asamizu, Kazusa DNA Research Institute, The First Laboratory for  
 RL Plant Gene Research; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 RL (E-mail:asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/,  
 RL Tel:81-438-52-3935(ex.2336), Fax:81-438-52-3934)  
 XX [2]  
 RN Asamizu E., Nakajima M., Kitade Y., Saga N., Nakamura Y., Tabata S.;  
 RA "COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS OF  
 RT PORPHYRA YEZOENSIS (RHODOPHYTA), BASED ON EXPRESSED SEQUENCE TAG FREQUENCY  
 RT ANALYSIS";  
 RL J. Phycol. 0:0-0(2003).  
 XX Key Location/Qualifiers

source 1. .489  
 /db\_xref="taxon:2788"  
 /mol\_type="mRNA"  
 /organism="Porphyra yezoensis"  
 /clone="PF004b08\_r"  
 /dev\_stage="sporophytes"  
 /strain="TU-1"  
 XX Sequence 489 BP; 100 A; 136 C; 170 G; 83 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 3.42e-20 Length: 489  
 Score: 264.00 Matches: 68  
 Percent Similarity: 57.14% Conservative: 24  
 Best Local Similarity: 42.24% Mismatches: 48  
 Query Match: 14.15% Indels: 22  
 DB: 6 Gaps: 5  
 US-09-921-992-50 (1-372) x AU186794 (1-489)  
 QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123  
 DB 14 TGTTTGAGA---AGCCCAAGAGGCGCGGAGTACTCGGACGAGAGTTCAACCGGC 70  
 QY 124 -----CysAlaArgAspLysAsn----- 129  
 DB 71 TCAAGACAGCTCAAGCCACCCTCACACCCTGTTGTCGCTCAAGAGCAGGGCA 130  
 QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLysTyTyr 149  
 DB 131 AGGCCATG-CGTATTGGGTCAACACACGGTCCCTGGCGGAGCGCATGTATGTTACCTAC 189  
 QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169  
 DB 190 GGGGAC---ACCCCGGGGGATGTCGAGTCCGGGATGAGTGCATTCAGATCTGCCCGG 246  
 QY 170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189  
 DB 247 GAGCTAGACTTTCACAACTCGTCTATTTCGATGAGGCGTCCCAAGGTGCCAGTCATGC 306  
 QY 190 GluSerTyArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHis 204  
 DB 307 GCAGCTACCGCCAGCTGGCGCTGGCGTCCGACGCTGAGGGGTACGACTACCCATCCAC 366  
 QY 205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224  
 DB 367 CTGGGTGTACTGAGCGCGGCGGAGTATGCGGGGTGAAAGAGCTCCATCGGTATC 426  
 QY 225 GlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244  
 DB 427 GGCACGCTCTTGGGGAGGGCATCGGCGACACGCTCGGGGTGAGCTTGACAGAACCC 486  
 QY 245 Val 245  
 DB 487 ATG 489  
 RESULT 14  
 ID AU042702  
 LOCUS  
 DEFINITION ST23E04 Pine Triplex shoot tip library Pinus taeda cDNA clone  
 ST23E04, mRNA sequence.  
 ACCESSION AU042702  
 VERSION AU042702.1 GI:5903147  
 KEYWORDS EST.  
 SOURCE Pinus taeda (loblolly pine)  
 ORGANISM Pinus taeda  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
 REFERENCE 1 (bases 1 to 601)  
 AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.  
 TITLE The Pine Gene Discovery Project  
 JOURNAL Unpublished

SOURCE  
ORGANISM

REFERENCE  
AUTHORS

TITLE

JOURNAL

COMMENT

Hordeum vulgare subsp. spontaneum  
Eukaryota; Viridiplantae; Streptophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;  
Triticeae; Hordeum.  
1 (bases 1 to 610)  
Sato.K., Saisho,D. and Takeda,K.  
Barley EST sequencing project in NIG and Okayama Univ  
Unpublished  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute Of Genetics  
1111 Yatae, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.

FEATURES  
source

/organism="Hordeum vulgare subsp. spontaneum"  
/mol\_type="mRNA"  
/strain="H602"  
/db\_xref="taxon:77009"  
/clone="bah58:01"  
/tissue\_type="top three leaves"  
/dev\_stage="adult, heading stage"  
/clone\_lib="K. Sato unpublished cDNA library, strain H602 adult, heading stage top three leaves"  
BASE COUNT 176 a 118 c 155 g 160 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: Length: 610  
Score: 269.50 Matches: 68  
Percent Similarity: 51.49% Conservative: 36  
Best Local Similarity: 33.66% Mismatches: 69  
Query Match: 14.44% Indels: 29  
DB: 12 Gaps: 5

US-09-921-992-50 (1-372) x BJ481080 (1-610)

QY 18 GlyAsnValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArg 37  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 10 GGGAACGTGGCACTCGGCAGTGATNCCCATGAGGATTTCAGACTATGACTACTCAGAT 69  
  
QY 38 ThrThraSpValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAasp 57  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 70 ACCAAGATGTTGCCAAGACTGTGGAGAGGTGATGAGGATTCGACAGAGCGGCTGAT 129  
  
QY 58 IleValArgValSerValProThrMetAspAlaAlaGluAlaPhe----- 72  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 130 TTGTTTAGAATAACC GTCCAGCTCCAGGGTAAAGGAAGCTGATCGCTGCTTGAGATTAAAGAAC 189  
  
QY 73 LysLeuIleLysGlnGlnValAsnValProLeuValalaAspIleHisPheAspTyrArg 92  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 190 ACTCTTTGTCCAGAAAGTAACAACATCCCTCTAGTGGCCGATATCCATTTTGGCGCTACA 249  
  
QY 93 IleAlaLeuLysValAlaGluTyrgLyValaAspCysLeuargileasnProGlyAsnIle 112  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 250 GTAGCTTTAGATGGTGCTGAATGC---TTTGACAAAATCCGTGTTAACCCAGGAACCTTT 306  
  
QY 113 GlyAsnGlu----- 115  
:::  
Db 307 GCCGATCGCGTCCCATTGTAAGAGCTGGAATATCTACGACGAGGATTACGAAAAAGGAG 366  
  
QY 116 ---GluArgIleArgMetVal-----valaspcysalaaargaspLysAsnIle 130  
||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 367 CTTGAACACATTCAGAGGGCTTTTCTCTCCATTTGGTTGAGAATGCCAAGAAGATGCGAAGA 426  
  
QY 131 ProIleargIleGlyValAsnAlaGlySerLeuGlulysaspLeuGlnGluLysTyrGly 150  
:||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
Db 427 GCCATCGGTATCGGAACAAATCATGGTAGTCTTTCTGACCCGGATATFAGACTACTACTGGT 486  
  
QY 151 GluProthrProGlnAlaLeuLeuGluSerAlaMetArghisValaspHisLeuAspArg 170

QY 48 IleLeuAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67  
 Db 358 GTATGAGGTAGCAGCATAAGGGCTGATTTGTTAGTAACATCCAGGCTAGAAAG 417  
 QY 68 AlaAlaGluAlaPhe-----LysLeuIleLysGlnValAsnValPro 82  
 Db 418 GAAGCTGATGCTGCTTTGAGATTAAGAACACTCTTTGTTGAGAAAGAAATTACAACTCC 477  
 QY 83 LeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyVal 102  
 Db 478 CTAGTGGCTGATTAATCATTTGCCCCGACAGTTGCTTTAGAGTGCTGAATGC---TTT 534  
 QY 103 AspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu----- 115  
 Db 535 GACAAATTCGTCAACCCAGGGAATTTGCTGATCGCGCTGCCCAATTTGAGCAGCTT 594  
 QY 116 -----GluArgIleArgMetVal----- 121  
 Db 595 GAATATACTGAAGATGATTATCAAAAGAGCTTTGAGCATATCGAAGGTCTTCTCCCG 654  
 QY 122 ---ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
 Db 655 TTGTTGAGAAATCGAAGCTAGTATGAAGCAATCGTATAGAACAAATCATGTGAGT 714  
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
 Db 715 CTGCTGACCGCATATGATGTTACTATGTTGAT---TCTCCAGCGGAATGTTGAGTCT 771  
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPhenAspGlnPhenLysValSerVal 180  
 Db 772 GCTTTGGAATTTTCAGGATCTGTGCG-AGCTGAGCTTCCATAACTTTGTTTTCATG 830  
 QY 181 LysAlaSerAsp 184  
 Db 831 AAGCAAGTAAC 842

## RESULT 10

CA022320 602 bp mRNA linear EST 23-OCT-2002  
 LOCUS HZ42M11r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ42M11  
 DEFINITION S-PRIME, mRNA sequence.

ACCESSION CA022320

VERSION CA022320.1 GI:24299694

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE Ruckelshaus, V., Zhang H., Weschke, W., Potokina, E. and Wobus, U.

AUTHORS Barley ESTs from developing seeds

TITLE Unpublished

COMMENT Contact: Stein Nils

Molecular Markers Group, Department Genbank

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3, 06466, Gatersleben, Germany

Tel: 039482-5522

Fax: 039482-5595

Email: stein@ipk-gatersleben.de

Insert length: 602 Std Error: 0.00

Plate: 42 row: M column: 11

Seq primer: M13rev.

Location/Qualifiers

1. 602

/organism="Hordeum vulgare subsp. vulgare"

/mol\_type="mRNA"

/cultivar="barke"

/db\_xref="GABI:275726"

/db\_xref="taxon:112509"

/clone="HZ42M11"

/tissue\_type="pericarp"

/dev\_stage="0-7 DAP (days after pollination)"

## FEATURES

source

RESULT 11

BJ481080

LOCUS

DEFINITION

CDNA clone bap58101 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

EST.

/lab host="XL10-Gold"

/clone lib="HZ"

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); pericarp 0-7 DAP(days after pollination). Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable. Average insert size is 900 bp"

BASE COUNT 169 a 110 c 156 g 167 t

ORIGIN

Alignment Scores:

Pred. No.: 7,14e-21 Length: 602  
 Score: 271.00 Matches: 70  
 Percent Similarity: 51.52% Conservative: 32  
 Best Local Similarity: 35.35% Mismatches: 62  
 Query Match: 14.52% Indels: 34  
 DB: 14 Gaps: 6

US-09-921-992-50 (1-372) x CA022320 (1-602)

QY 55 GlyAlaAspIleValArgValSerValProThrMetAspAlaAlaGluAlaPhe----- 72

Db 15 GCGCGTGAATTTGTTAGATAACCGTCCAGGCTAAAGAGAGCTGATCGCTGCTTTGAG 74

QY 73 -----LysLeuIleLysGlnValAsnValProLeuValAlaAspIleHisPhe 89

Db 75 ATTAAGAACACTCTTGTCCAGAGAAATTACACATCCCTCTAGCGGCCGATATCCATTTT 134

QY 90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109

Db 135 GCGCTACAGTAGCTTTAGAGTGGCTGATGC---TTTGACAAAATCCGTGTTAACCA 191

QY 110 GlyAsnIleGlyAsnGlu----- 115

Db 192 GGAACTTTGCGGATCGCGCTGCCCAATTTGAAAAGCTGGAATATATCTGAAGACGATTAC 251

QY 116 -----GluArgIleArgMetVal-----ValAspCysAlaArgAsp 127

Db 252 GAAAGAGAGCTTGAACACATTCAGAGGGTCTTTTCTCCATTTGTTGAGAAATGCAAGAAG 311

QY 128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147

Db 312 TATGGAAGAGCCTGCGTATCGAACAAATCATGTAGTCTTTCTGACCGGATATGAGC 371

QY 148 LysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHis 167

Db 372 TACTATGGTGAT---TCTCAAGGGGAATGGTTGAGTCTGCTTTGGAATTTGCTAGGATC 428

QY 168 LeuAspArgLeuAsnPhenAspGlnPhenLysValSerValLysAlaSerAspValPheLeu 187

Db 429 TGTCCGAATTTGCGACTTCCATACTTTGATTTTCAATGAAAGCAAGTAACCTGTTGTC 488

QY 188 AlaValGluSerTyrArgLeuLeuAlaLysGlnIle-----AspGlnPro 202

Db 489 ATGGTCCAGCATATCGCTGCTTTGAGCGGAATGTATTAACCTTGGATGGGATATCTCT 548

QY 203 LeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220

Db 549 TTGCACCTTGGAGTTACCGAAGCTGGTGGGGTGAAGATGGGAGGATGAAGTCT 602

BJ481080 610 bp mRNA linear EST 23-MAY-2002

BJ481080 K. Sato unpublished cDNA library, strain H602 adult,

heading stage top three leaves Hordeum vulgare subsp. spontaneum

CDNA clone bap58101 5', mRNA sequence.

ACCESSION BJ481080

VERSION BJ481080.1 GI:21159548

KEYWORDS EST.

```
/db_xref="taxon:5855"  
/dev_stage="asexual blood forms"  
/lab_host="Saimiri boliviensis"  
/clone_lib="Pv MBN #30"  
/notes="Vector: pBluescript SK(+) vector DNA, phagemid  
excised from lambda ZAP; Site 1: EcoR V; Site 2: EcoR V;  
Host leukocytes were extracted from P. vivax infected  
blood using the following methods: first, infected blood  
was activated by the addition of 0.5 ml of ADP (40mg/ml)  
per 10 ml blood. Then blood was passed over a column of  
acid washed 0.1 mm glass beads, then through a Plasmidipur  
filter, followed by passage through a column of pre-wet  
Whatman Cfil powder (1:2 ratio volume of blood to Cfil),  
and finally centrifuged through a 50% Percoll density  
cushion. Purified DNA was digested with mung bean nuclease  
in the presence of 4% formamide at 50°C as described  
(Vernick, K.D., ImberSKI, R.B., and McCutchan, T.F. 1988.  
Nucleic Acids Research 16:6883-6896). Digested DNA was  
blunt-ended using T4 DNA Polymerase and size fractionated  
over a Sepharose CL-2B column. Fractions in the size range  
500bp-4kb were ligated into the Eco RV site of pBluescript  
SK(+), and E. coli XL-10 Gold transformed with the  
ligation mixture."
```

BASE COUNT 178 a 150 c 119 g 206 t 5 others  
ORIGIN

Alignment Scores:  
Pred. No.: 5,01e-25 Length: 658  
Score: 307.50 Matches: 79  
Percent Similarity: 53.85% Conservative: 33  
Best Local Similarity: 37.98% Mismatches: 61  
Query Match: 16.48% Indels: 35  
DB: 28 Gaps: 6

US-09-921-992-50 (1-372) x AZ570993 (1-658)

```
QY 73 LysLeuIleLysGlnValAenValProLeuValAlaAspIleHisPheAspTyrArg 92  
DB 642 AAAATGGAGCTCAAAATATTACATTCCTGTTAGTAGACATACATTTATCCCAAA 583  
QY 93 IleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAenIle 112  
DB 582 ATTCTCTCATGCAAGCAGATGTA---TTTGATAAAATACGCATACCCAGCGAATTAC 526  
QY 113 GlyAenGluGluArg----- 117  
DB 525 GTCGATGGGAGNAGAAATGGATCAATAAGATTACAAACGAGAGAGANTTCGATCAA 466  
QY 118 -----IleArgMetValValAspCysAlaArgAspLys 128  
DB 465 GGGAAATTATTATTCAGNAGAGTTTCATCCCTTGATTGAAAGTGAAGG---TTA 409  
QY 129 AenIleProIleArgIleGlyValAenAlaGlySerLeuGluLysAspLeuGlnGluLys 148  
DB 408 AATAGAGCCATACGAATCGGCAAAATCACGGCTCCTATCTTNTAGGCTACTCTCTTC 349  
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168  
DB 348 TATGGGGAT---ACCCNNTAGGTATGATGTCGAATCGCGCTTTTCAATTTTCCCGATTGTG 292  
QY 169 AspArgLeuAenPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188  
DB 291 GTACAGATAACTTCTTCATGTGGTTTCTCATGAAGGCTTCCACCGGTATATATG 232  
QY 189 ValGluSerTyrArgLeuLeu---AlaLysGlnIleAspGln----- 201  
DB 231 ATACAGTCGTATAGGCTACTCGTGGCGAGGCAATATGAAAGGATGGGCAACGGTTACTA 172  
QY 202 ---ProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220  
DB 171 TTCCCACTGCATTTGGCGCTTACCGAAGCAGGGTTTGGGGATAACGGAAGGATTAATCC 112  
QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
```

```
Db 111 TATTTGGGATAGGTTGCTGCTGTACGAGGGATAGGACACCATCAGGATTCGTGA 52  
QY 241 AlaAlaAspProValGluGluIle 248  
Db 51 ACTGAGGACCCCTTGGGAGGAATTA 28  
RESULT 9  
LOCUS CB628479 852 bp mRNA linear EST 08-APR-2003  
DEFINITION OSIIIEB04B06.f OSIIIEB Oryza sativa (indica cultivar-group) cDNA  
clone OSIIIEB04B06 5', mRNA sequence.  
ACCESSION CB628479  
VERSION CB628479.1 GI:29623468  
KEYWORDS EST.  
SOURCE Oryza sativa (indica cultivar-group)  
ORGANISM Oryza sativa (indica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzeae; Oryza.  
REFERENCE 1 (bases 1 to 852)  
AUTHORS Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
TITLE Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
JOURNAL Unpublished  
COMMENT Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://Genome.arizona.edu  
PCR Primers  
FORWARD: gta aaa cga cga cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 04 row: B column: 06  
Seq primer: gta aaa cga cga cca gtcg.  
FEATURES  
source  
1..852  
/organism="Oryza sativa (indica cultivar-group)"  
/mol_type="mRNA"  
/cultivar="IR36"  
/db_xref="taxon:39946"  
/clone="OSIIIEB04B06"  
/tissue_type="Leaf"  
/dev_stage="3 week"  
/lab_host="DH10B"  
/clone_lib="OSIIIEB"  
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2:  
XhoI; 24 hrs after inoculation with Rice Blast (P06-6-3)"  
BASE COUNT 233 a 172 c 223 g 224 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.69e-21 Length: 852  
Score: 276.50 Matches: 72  
Percent Similarity: 53.43% Conservative: 37  
Best Local Similarity: 35.29% Mismatches: 66  
Query Match: 14.82% Indels: 30  
DB: 14 Gaps: 5  
US-09-921-992-50 (1-372) x CB628479 (1-852)  
QY 8 GlnArgArgLysSerThrArgIleTyrValGlyAenValProIleGlyAspGlyAlaPro 27  
DB 238 AGGAGAGAAAAACCCGACCTGTGATGTTGGGATGTGCCACTTGGCAGTCATATCCC 297  
QY 28 IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAenGln 47  
DB 298 ATTAGGATTCAGATATGACCACCTCGGATACCAAGGATGTTGCTAAAAACCGTAGGAG 357
```

```

Db 71 TGC---TTTGACAAAATACGTGTCAATCCTGGAAACCTTCTGCACAGCGGAGCCAGTTT 127
QY 115 -----GluGluArg 117
Db 128 GAGCAATTAGTACACAGAGATGACTATCAGAAAGAACTCGAGCATATTGAGGAGTT 187
QY 118 IleArgMetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsn 137
Db 188 TTTACACCATGTGTGGAATAATGTAAGAGATGAGCGTCCATCGCATTTGGACAAAC 247
QY 138 AlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeu 157
Db 248 CATGGGAGCCCTTCAGATCCATTATCAGCTATTATGGGAC---TCGCTAGGGAATG 304
QY 158 LeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLys 177
Db 305 GTAGATACGACATTTGAGTTTGCAGAAATTTGTAAGAAAGTTGACATTTCCAAATTTTCGTC 364
QY 178 ValSerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLys 197
Db 365 TTCTCAATGAAGTAGCAATCCAGTAGTATGTTTTCAGCGTATCGCTTCTCTAGCT 424
QY 198 GlnIle-----AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGly 212
Db 425 GAGATGTATGTTTCAGGATGGGACTATCCTTTACACTTGGAGATTACTGAAGCTGCTGAA 484
QY 213 AlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSerGluGlyIle 232
Db 485 GGTGAGGTGGAGGATGAAGTCTCAATCGGTATTGGCACACTTCTTCAGGATGTTTG 544
QY 233 GlyAspThrLeuArgValSerLeu 240
Db 545 GGTGATACATACGGGTTTCCCTA 568

```

```

RESULT 7
BH235010
LOCUS BH235010 720 bp DNA linear GSS 01-JAN-2002
DEFINITION PH_05.x Spiroplasma kunkelii H Spiroplasma kunkelii genomic clone
ACCESSION PH_05.x, genomic survey sequence.
VERSION BH235010
KEYWORDS BH235010 GI:18030478
SOURCE GSS.
ORGANISM Spiroplasma kunkelii
Bacteria; Firmicutes; Mollicutes; Entomoplasmatiales;
Spiroplasmataceae; Spiroplasma.
REFERENCE Hogenhout, S.A.
AUTHORS Genomic sequences from Spiroplasma kunkelii strain M2
TITLE Unpublished
JOURNAL
COMMENT Contact: Hogenhout SA
Department of Entomology
The Ohio State University-OARDC
120 Thorne Hall, 1680 Madison Avenue, Wooster, OH 44691, USA
Tel: 330 263 3730
Fax: 330 263 3686
Email: hogenhout.1@osu.edu
Class: HindIII fragments.
Location/Qualifiers
1. .720
/organism="Spiroplasma kunkelii"
/mol_type="genomic DNA"
/strain="M2"
/db_xref="taxon:47834"
/clone="PH_05.x"
/clone_lib="Spiroplasma kunkelii H"
BASE COUNT 255 a 96 c 132 g 233 t 4 others
ORIGIN

```

```

Alignment Scores: 1.32e-25 Length: 720
Pred. No.: 313.00 Matches: 67
Score:

```

```

Percent Similarity: 62.20% Conservative: 35
Best Local Similarity: 40.85% Mismatches: 61
Query Match: 16.77% Indels: 2
DB: 28 Gaps: 0
US-09-921-992-50 (1-372) x BH235010 (1-720)
QY 190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGlu 209
Db 31 CAAGCTTATACCTTAGCTAGTAAGAATTAATCCCTTCATCTAGGGATGCT-GAA 89
QY 210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSer 229
Db 90 GCTGTGTAGTCATCATACCGAAACAAANNAATCATGTAGTGTGTGTGTCACCGCTTCTTTT 149
QY 230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLys 249
Db 150 AATGTATTGGTGTAGTACGATCCGAATTTATCAACTGTATCAATCAGCAGGTGGAA 209
QY 250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269
Db 210 GTTGTAAACGAATGTTAAATTCATTAGGACTTTATGATAACATTTGTCATATAATGCT 269
QY 270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289
Db 270 TGCCCAACATGTGTGTCGTTAGATATGACCTTTTCCCGCTTGTAAAGAAATTAAGAAGAA 329
QY 290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309
Db 330 TATACAAAAGATTTAAAAATTCATTAATAATTCATTAATAATTCATTAATAATTCATTAAT 389
QY 310 -ProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGlyLe 329
Db 390 CCCAGAGNAGCTAAACAGCTGATTAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 449
QY 329 uTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeuG 349
Db 450 TTTTAAAAAGGTAAAAATTTATAAATCATGTATAAACAAGAAGATCTTTGTTCCAGAAATA 509
QY 349 uAlaArgIle 352
Db 510 ACTATTAATT 519
RESULT 8
AZ570993/c
LOCUS AZ570993 658 bp DNA linear GSS 15-MAY-2001
DEFINITION 281PVD07 Pv MEN #30 Plasmodium vivax genomic 3', genomic survey
ACCESSION AZ570993
VERSION AZ570993
KEYWORDS GSS.
SOURCE Plasmodium vivax (malaria parasite P. vivax)
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 658)
AUTHORS Carlton, J.M.-R. and Dame, J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13 (-20) forward
Class: shotgun.
Location/Qualifiers
1..658
/organism="Plasmodium vivax"
/mol_type="genomic DNA"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"
FEATURES
source

```





```

QY 247 GluIle 248
Db 602 GAGATT 607

RESULT 4
AY104363
LOCUS Zea mays PCO116062 mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO116062 mRNA sequence.
ACCESSION AY104363
VERSION AY104363.1 GI:21207441
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1442)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whittitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1442)
Coe, E.H.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
If you are interested in getting corresponding physical clones,
these are publicly available from ZmDB and may be found by BLAST
searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schnable, Iowa State, then clones may be requested from ZmDB:
www.zmdb.iastate.edu.
LOCATION/Qualifiers
1..1442
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:636753"
/db_xref="taxon:4577"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/notes="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"

BASE COUNT 381 a 282 c 374 g 399 t 6 others
ORIGIN

Alignment Scores:
Pred. No.: 9,68e-27 Length: 1442
Score: 326.50 Matches: 99
Percent Similarity: 49.66% Conservative: 46
Best Local Similarity: 33.90% Mismatches: 77
Query Match: 17.50% Indels: 71
DB: 11 Gaps: 11

US-09-921-992-50 (1-372) x AY104363 (1-1442)

QY 65 ThrMetAspAlaGluAlaPheLys-----LeuIleLysGlnValAsn 80
Db 3 ACAGGGAAGCTGATGCTCTTGAGATCAAGAACACTGTGGTTCAAGAAATTCAAC 62
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 63 ATTCCACTAGTGGCCGATATTCATTTTGTCTTCCACGCTAGCCCTAAGGGTGGCAGATGT 122
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu----- 115
Db 123 ---TTTGACAAAATTCGTGTGACCCAGGAAATTTTGTGTGATCTGCTGCTCAATTGAA 179

116 -----GluArgIleArgMetVal--- 121
180 AAGCTGGAATATACTGACGACGACTACCAAAAGAGCTAGACATATTGAGAAGGTGTTT 239
122 -----ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAla 138
240 TCTCCATTAGTTGAGAAATGCAAGCAGTAGTATGAAGAGCAATGCAATAGCAACAAATCAT 299
139 GlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeu 158
300 GGTAGTCTTTTCGACCGCATATAGCTACTATGGTGAT---TCTCCAGGGGAATGGTT 356
159 GluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysVal 178
357 GAGTCTGCTTGGAAATTTGCTAGATATATGTCGAAGTTGCACTTCCACAACATTGTTATT 416
179 SerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGln 198
417 TCGATGAAAGCTAGTAACCCCTGTATCATGTGTCACAGCATATCGCTGCTTGTGGCAGAA 476
199 Ile-----AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAla 213
477 ATGTATACCTAGGATGGGATTTATCCCTTGCACCTGGGTGTACAGAGCTGGAGGGGT 536
214 ArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGly 233
537 GAAGATGA-AGGATGAAATCTGCTATTGGCAATGGGACACTGCTAATGATGCTTTGGGT 595
234 AspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle----- 248
596 GATACAATCCGTCTCTCCCTCAGAACACCAGAGAGAGATTCATCCTTCCCAAGG 655
249 -----LysValGlyPheAspIleLeuLysSe 257
656 TTGGCAAAATCTTGGGACGCGAGCGCCAAACCTTCAAAATTTGGGTGGGCCCATTTGAAG-- 713
257 rLeuArgIleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnG 277
714 -----AAAAGCACAGGCGTTATTTGATTCCAGC----- 743
277 uPheAspValIleGlyThrValAsnAlaLeu---GluGlnArgLeuGluAspIleIleTh 296
744 -----GTAGGAGTGTCAATTCCTTTCAGAGAGGAGGTGAGGAAGTTGACTAC 793
296 rProMetAspValSerIleGlyCysValVal 307
794 AGAAATG-----TCTGTCATCGTG 812

RESULT 5
BZ569685/c
LOCUS BZ569685 1268 bp DNA linear GSS 17-DEC-2002
DEFINITION pac82-164_970.s1 pac82-164 Pseudomonas aeruginosa genomic clone
pac82-164_970, genomic survey sequence.
ACCESSION BZ569685
VERSION BZ569685.1 GI:27204646
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1268)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol., (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 23, 2003, 17:20:49 ; Search time 2600 Seconds  
(without alignments)  
3477.410 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAPIQRKRSRIYGVN.....RAKASQLDEARRIDVQVEK 372

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09921992/runat\_21112003\_162546\_18418/app\_query.fasta\_1.519  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09921992.@CGN\_1\_1\_2810@runat\_21112003\_162546\_18418 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_htc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_htc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pin.\*  
20: em\_gss\_vit.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	820.5	44.0	1323	29	BZ575759	BZ575759 meh2_4614	
2	379	20.3	828	14	CB669733	CB669733 OSJNE02H	
3	348.5	18.7	853	14	CB619369	CB619369 OSIIa030	
4	326.5	17.5	1442	11	AY104363	AY104363 Zea mays	
c	5	321.5	17.2	1268	29	BZ569685	BZ569685 pacs2-164
	6	315.5	16.9	579	12	B1920888	B1920888 EST540823
7	313	16.8	720	28	BH235010	BH235010 PH_05.X.S	
c	8	307.5	16.5	658	28	AZ570993	AZ570993 281PvD07
	9	276.5	14.8	852	14	CB628479	CB628479 OSIIb04B
10	271	14.5	602	14	CA022320	CA022320 H242M11r	
11	269.5	14.4	610	12	BJ481080	BJ481080 BJ481080	
12	265	14.2	706	10	BG591263	BG591263 EST499105	
13	264	14.1	489	6	AU186794	AU186794 Porphyra	
14	262	14.0	601	9	AM042702	AM042702 ST23E04.P	
15	261	14.0	603	12	BJ551332	BJ551332 BJ551332	
16	260	13.9	558	9	AV917069	AV917069 AV917069	
17	260	13.9	608	12	BJ465443	BJ465443 BJ465443	
18	260	13.9	616	9	AV934187	AV934187 AV934187	
19	259	13.9	484	13	BQ080994	BQ080994 san12Q09.	
20	259	13.9	651	9	AU252386	AU252386 AU252386	
21	257	13.8	640	13	BU091269	BU091269 6t70q12.Y	
22	252	13.5	579	28	AZ523878	AZ523878 224PB02	
c	23	250.5	13.4	351	13	BU004160	BU004160 QGG37N12.
	24	244	13.1	465	9	AI482869	AI482869 EST242192
25	236	12.6	576	13	BQ975044	BQ975044 QHI18D05.	
26	232.5	12.5	545	9	AV626844	AV626844 AV626844	
27	229	12.3	532	10	BE924238	BE924238 EST428007	
28	229	12.3	628	13	BU048059	BU048059 PP_LEA003	
29	229	12.3	680	12	BG890681	BG890681 EST516532	
30	229	12.3	700	12	B1933447	B1933447 EST553336	
31	229	12.3	728	10	BG525868	BG525868 53-47-1-2	
32	228	12.2	423	13	BQ862800	BQ862800 QG22B19.	
33	225	12.1	792	14	CB620208	CB620208 OSIIa05D	
34	225	12.1	821	14	CB291771	CB291771 UCR001.0	
35	225	12.1	821	14	CB628414	CB628414 OSIIb03P	
36	225	12.1	1127	29	BZ577824	BZ577824 msh2_5589	
37	224	12.0	687	14	CA781720	CA781720 046F07AF	
38	223.5	12.0	413	12	BM059608	BM059608 KS01002E0	
39	222.5	11.9	544	13	BQ591926	BQ591926 E012583-0	
40	222.5	11.9	731	10	BG124857	BG124857 EST470503	
41	222	11.9	640	9	AW738440	AW738440 EST339867	
42	218.5	11.7	770	14	CB662235	CB662235 OSJNE06B	
43	216	11.6	448	9	AV644003	AV644003 AV644003	
44	216	11.6	482	9	AV642898	AV642898 AV642898	
45	216	11.6	502	9	AV626792	AV626792 AV626792	

ALIGNMENTS

RESULT	1	1323 bp	DNA	linear	GSS 17-DEC-2002
BZ575759	msh2_4614.x1 msh	Pseudomonas aeruginosa	genomic clone msh2_4614,		
LOCUS	genomic survey sequence.				
DEFINITION	BZ575759				
ACCESSION	BZ575759.1	GI:27210820			
VERSION	GSS.				
KEYWORDS	Pseudomonas aeruginosa				
SOURCE	Pseudomonas aeruginosa				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
REFERENCE	1 (bases 1 to 1323)				

```

Db      597 ATAGAGCTTACAGGATGCTCGCCCTCTGTGATCTATCCTTCCATTGGGGTTACG 656
QY      209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db      657 GAGGGGGGAATCTTTTATAGCTCCAGATCAAAATCGCTATGGCTTTAGGGGGCTTTTA 716
QY      229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db      717 ATGGAGGGCATTTGGGATACGATGCGCTATCCATCACAGGGGAATTAGAAATGAATC 776
QY      249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPhelle 268
Db      777 AAAGTGGCCAGAGCAATTTTAGCCCATAGCGCGGCTTGAAGAAGAGGATTAATTGGATT 836
QY      269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db      837 TCTTGCCCCACCTGCGGGCGCATTTGAAGCCAATTTAGTGGATATGCGCATCAAGGTAGAA 896
QY      289 GlnArgLeuGluAspIleIleThrPrometAspValSerIleIleGlyCysValValAsn 308
Db      897 AAACGCTTAAGCCACATCAAAACCCCTTTAGACATTAGCGTGATGGGTTGCGTGGTGAAT 956
QY      309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyClyAsnLysLysSerGly 328
Db      957 GCTTTGGGTGAAGCCAAAGCATGCAGACATGCGCATCGCTTTTGGGAATCCGACGCGTTG 1016
QY      329 LeuTyrGluAspClyValArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
Db      1017 ATCATTAAGAGGGGTAAAGTCATTCAAACTGGCTGAAAGAGATTATTGAAACTTTT 1076
QY      349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
Db      1077 GTCATAGAGTGAATAATTAGCTAAAGAAAGAGAAAAAAGT 1118

```

Search completed: November 23, 2003, 17:55:54  
 Job time : 2022 secs









Qy	29	AlaValGlnSerMetThrAsnThrArgThrThraSpValGluAlaThrValAsnGlnIle	48
Db	3183	ACTATTCAAAGCATGACTACTACAAAGACACATGATGTCGAAGCAACAGTAGCGGAATT	3242
Qy	49	LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla	68
Db	3243	CACGGATTAGAAGAGTGGGTGTCAGATTGCGGAGTTGCTTGCTCTGATGACGCTGCA	3302
Qy	69	AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis	88
Db	3303	GCCAAATGCTCTTAGTGCCATCAAGAAAGAGTTCATATTCCTCTTGTCGAGATATTCAT	3362
Qy	89	PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn	108
Db	3363	TTTGATTACCGCATAGCACCTTAAAGCTATTGATGCAGGTGTGACAAAATCCCGGATTAA	3422
Qy	109	ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys	128
Db	3423	CCTGGTAACATCGTCCCGGTGATCGGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA	3482
Qy	129	AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys	148
Db	3483	AATATTCCAATCCGATTGGGGTTAATGCTGGTAGTTTAGAAAAAGAAAATATTTCAAAAA	3542
Qy	149	TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu	168
Db	3543	TACGGTTACCCCTACTGCTGAAGGAATGGTAGAAGTGCACATTGATCATATTAATAATTC	3602
Qy	169	AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla	188
Db	3603	GAAGATTAGATTTTATGATATCATTTGTTCTTTGAAGCTTCTGATGTGAATTTAGCA	3662
Qy	189	ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr	208
Db	3663	ATTGAAGCTTATGATAAAGGTAGTCGCGCATTTAATATCTCTPACATCTTGGAAATPACA	3722
Qy	209	GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu	228
Db	3723	GAATCTGGTAGCGAGTTGCTGGAGGATAAAAAGTGTCTGCTGTTTAGAGCGCATCTC	3782
Qy	229	SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle	248
Db	3783	AGTTTGGGCATTGGAATACATTACGGGTATCTTTGAGTGTGATCTGTGGGAAGAGATA	3842
Qy	249	LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle	268
Db	3843	AAAGTGCCCGGAGGTTTTAAATTCGTTTGGCCCTTTCCTCGAATCCGCCATGCTTATC	3902
Qy	269	AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu	288
Db	3903	TCCTGCCCTACTTCCGTCGNATAGAGATTGATTGATTTCGTATCGCTAATGAAGTGAA	3962
Qy	289	GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn	308
Db	3963	AATTACATAGCAACGATTAAAGCTCCGATTAAGTAGCCGTGCTGTCGCGTCAAC	4022
Qy	309	GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly	328
Db	4023	GGTCCCGGAAGAGCTCGCGAAGCAGATATCGGAATTCGTGGTCCAAATGGAGAAGGCC	4082
Qy	329	LeuTyrGluAspGlyValAlaGlyAspArgLeuAspAsnAsnAspMetIleAspGlnLeu	348
Db	4083	CTTTTATAGCATGTTAAATCATCCGAAAGTACAGAGCCATTATGGTAGAAGACTT	4142
Qy	349	GluAlaArgIleArgAlaLysAlaSerGlnLeu	359
Db	4143	AAGAAAGGAATTTGATATTTTGGCCAGGAAGATA	4175

XX	13-AUG-2002	(first entry)
DT		
XX		
DE	Bacillus licheniformis	genomic sequence tag (GST) #1692.
XX		
KW	Differential gene expression;	genomic sequenced tag; GST;
KW	altered culture condition;	environmental stress;
KW	physiological provocation;	ds.
OS	Bacillus licheniformis.	
XX		
XX	WO200229113-A2.	
PN		
XX		
PD	11-APR-2002.	
XX		
PF	05-OCT-2001; 2001WO-US31437.	
XX		
PR	06-OCT-2000; 2000US-0680598.	
PR	27-MAR-2001; 2001US-279526P.	
XX		
XX	(NOVO ) NOVOZYMES BIOTECH INC.	
PA	(NOVO ) NOVOZYMES AS.	
XX		
PI	Berka R, Clausen IG;	
XX		
DR	WPI; 2002-416684/44.	
XX		
PT	Monitoring differential expression of several genes in fi	
PT	cell relative to expression of same genes in one or more	
PT	Bacillus cells, by using substrate containing Bacillus ge	
PT	sequenced tag array	
XX		
PS	Claim 4; SEQ ID NO 1692; 200pp; English.	
XX		
CC	The invention describes a method of monitoring differenti	
CC	genes in a first Bacillus cell relative to expression of	
CC	other Bacillus cells, comprising hybridising labelled nuc	
CC	isolated from Bacillus cells to a substrate containing ar	
CC	genomic sequenced tags (GST), examining the array, and de	
CC	relative gene expression by an observed hybridisation rep	
CC	a spot in the array. The method is useful for measuring t	
CC	genes in a first Bacillus cell relative to expression of	
CC	in one or more second Bacillus cells. The method is usefu	
CC	Global expression of several genes from a Bacillus cell,	
CC	genes, identifying possible functions of unknown open rea	
CC	monitoring gene copy number variation and stability. Moni	
CC	in expression of genes may be used to provide a represent	
CC	in which Bacillus cells adapt to changes in culture condi	
CC	environmental stress or other physiological provocation.	
CC	follow-up characterisation is unnecessary, when one spot	
CC	equals one gene or one open reading frame, since sequenc	
CC	available. This sequence represents a genomic sequence ta	
CC	the method of the invention.	
CC	Note: The sequence data for this patent did not form part	
CC	specification, but was obtained in electronic format dire	
CC	at	
CC	ftp.wipo.int/pub/published pct sequences.	

```
Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2
US-09-921-992-50 (1-372) x ABA03041 (1-2944528)

QY 10 ArgLysSerThrArg---lleTyrValGlyAsnValProIleGlyAspGlyAlaProfile 28
Db 1474820 CGCGAAACACATCGCCAGTCCCAAGTGGGTAAATTAATCTATTGGTGGTAGGAGAAATTA 1474879

QY 29 AlaValGlnSerMetThrSerThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db 1474880 ACTATCCAAGACATGACTACTACAAAGACACATGATGTGCGAAGACACATGACGAGAAAT 1474939

QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 1474940 CACCGATTAGAGAGAGCTGGTGTCTAGATGTGCGAGTGTCTGTCTCTGATGACGTGCA 1474999

QY 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88
Db 1475000 GCGAATGCTCTTAGTCCCATCAAGAAAAAGATTCAATTCGCTTGTGCGACAGATATTCAT 1475059

QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db 1475060 TTGTGATTACCGACTAGCACTTAAGCTATTGATGCGAGGTGTTGACAAAAATTCGGATTAT 1475119

QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db 1475120 CTGTGTAACTATGTCGCCGTGATCGGTGGGAAAGTGGTAAATGCTGCTTAAGCAAAA 1475179

QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db 1475180 AATATTCCAATCCGTATGCGGTAAATGCTGTGTAGTTAGTAAAGAAAAATTTATCAAAA 1475239

QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 1475240 TATGTTACCTACTGCTGACCGAATGGTAGAAGTGCATCTGCCCATATTAATAATTCCTC 1475299

QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 1475300 GAAGATTAGATTTTATCATATCATCTTTTCTTTGAAGCTTCTGATGTGAATTTAGCA 1475359

QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 1475360 ATTGAAGCTTATGATAAGCTAGTCGCGCATTAATTAATCTCTGATCTCGGAATTACA 1475419

QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 1475420 GAATCTGGTACACAAATTTCTGCGAGCAATAAAAGTGTCTGCTGGTTTAGGAGCGGATCTC 1475479

QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db 1475480 AGTTTGGGCAATTTGGAATAATCATACAGATATCTTTGAGTGCTGATCCTGTGGAAGAAATPA 1475539

QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 1475540 AAAGTGGCCGCGAAGTTTAAATCATTTGCTCTTCTCCAAATGCCCATGCTTATC 1475599

QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 1475600 TCCTGCCCTACTTGGGTGCGATAGATGATTGATTAAATTCGTATCGCTAATGAAGTGGAA 1475659

QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 1475660 AATTACATAGCAAAAGATTGAAGTTCGATTAAGTAGCGCTGCTGGCTGCGGTCAAC 1475719

QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyAsnLysLysSerGly 328
Db 1475720 GGCCCTGGAGAGCTCGCAAGCCGATATCGGAATTCGTGGTTCACACGGAGAGCGCCTT 1475779

QY 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
Db 1475780 CTTTTTAGACATGGTAAAAATTTACGAAAAGTACCGGAAGCTATTATGATGACGAACCTT 1475839
```

```
US-09-921-992-50 (1-372) x ABO69993 (1-1107)
QY 10 ArgLysSerThrArg---IleTyrValGlyAenValProIleGlyAspGlyAlaProIle 28
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 78
19 CGCGAAACACCTCGCCAGCTCGCAAGTGGGTAATTAACCTATTGGTGGTGGTGGTGGT 78
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 138
79 ACTATCCAAGCATGACTACTCAAGACACATGATGTCGAAGCAACACAGTACGCAAAAT 138
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 198
139 CACCGATTAGAGAAGCTGGTGTCTCAGATTGGCGAGTTGCTGTCTGATGAACGTGCA 198
QY 69 AlaGluAlaPheLysLeuLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 258
199 GCGAATGCTCTTAGTGCCATCAAGAAAGATTCATATTCGCTTGGCGAGATTCAT 258
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 318
259 TTGTATTACCGACTAGACTTAAAGCTATTGATGCGAGGTGTGACAAATTCGATTAT 318
QY 109 ProGlyAsnIleGlyAsnGluAlaGlyIleArgMetValValAspCysAlaArgAspLys 128
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 378
319 CTGTGTAACTATGGTCCGCGTATCGGTGGAAAGTGTGTTAATGCTGCTAAAGCAAAA 378
QY 129 AsnIleProIleArgIleGlyValAlaAlaGlySerLeuGluLysAspLeuGluLys 148
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 438
379 AATATTCCAATCCGATTATGGGTTAATGCTGTGGTAGTTTGTAGAAAGAAATTTATCA 438
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 498
439 TATGGTATCCCTACTGCTGACGAATGGTAGAAGTGCACCTTGCCCATATTAATTCCTC 498
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 558
499 GAAGATTATAGATTTTATGATATCATTTGTTCTTCAAGGCTTCTGATGTGAATTAGCA 558
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 618
559 ATTGAAGCTTATGATAAGCTAGTCGCGCATTTAATATATCTCTGCTGCAATTTACA 618
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 678
619 GAATCTGGTACAAATTTCTGTGAGGAATAAAGTGTCTGTTAGAGCGGATCTACTC 678
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 738
679 AGTTTGGGCATTTGGAATACATTTACAGTATCTTTGAGTGTCTGCTGTGGAAGAAATA 738
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 798
739 AAAGTGGCCGGGAGAGTTTAAATCATTTGTTCTCTCGAATGCCGCAATGCTTATC 798
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 858
799 TCTGCGCTTACTTGGCGTGAATAGATGATTAAATTCGTATCGTATCGTAAAGTGGAA 858
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleGlyCysValValAsn 308
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 918
859 AATTACATAGCAAGATTGAAGTTCGATTAAGTAGCCGCTGCTGCTGCGGTCAAC 918
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyAsnLysLysSerGly 328
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 978
919 GGCCCTGGAGAGCTCGCAAGCCGATATCGGAATTCGTTGTTCAACGAGGAGGCGCTT 978
QY 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1038
979 CTTTTTAGACATGGTAAATTTATTCGAAAGATGACCGAAGCTATTATGATAGCAACTT 1038
QY 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
DB |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 1107
```

```
DB 1039 AAGAAAGAAATGATATTTTGGCAGAGAATTT---TTTGTGAGAAATAGATTGGAA 1095
QY 369 GlnVal 370
DB 1096 AGCCTT 1101
RESULT 10
ABA03041
ID ABA03041 standard; DNA; 2944528 BP.
XX ABA03041;
XX 05-FEB-2002 (first entry)
XX Listeria monocytogenes EGD-e genome sequence.
DE Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW vitamin B12; bacterial infection; disease; ds.
XX Listeria monocytogenes.
OS WO200177335-A2.
PN 18-OCT-2001.
XX 11-APR-2001; 2001WO-FR01118.
XX 11-APR-2000; 2000FR-0004629.
XX (INSP ) INST PASTEUR.
PI Buchrieser C, Frangoul L, Couve E, Ruhnok C, Fsihi H, Dehoux P;
PI Dusurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT and prevention of Listeria and related bacterial infections, and
PT related polypeptides
XX Claim 1; SEQ ID No 1; 192pp; French.
XX The present sequence is the genome sequence of Listeria monocytogenes
CC EGD-e. This sequence and fragments of this sequence are useful for
CC selecting probes and primers for detecting genes in L. monocytogenes and
CC related organisms, and to study genetic polymorphisms and other genomes.
CC Proteins (ABBA7297-ABBA50149) expressed from the present sequence are
CC useful for raising specific antibodies, identification of L.
CC monocytogenes, and related organisms, and for biosynthesis and
CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and
CC proteins encoded by it are also useful for selecting compounds that
CC regulate gene expression and cell replication and modulate L.
CC monocytogenes-related diseases. In addition, this sequence and proteins
CC encoded by it are useful in pharmaceutical and vaccines compositions for
CC the treatment or prevention of infections by L. monocytogenes and related
CC organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;
Alignment Scores:
Pred. No.: 9,72e-76 Length: 2944528
Score: 868.00 Matches: 178
Percent Similarity: 67.68% Conservative: 67
Best Local Similarity: 49.17% Mismatches: 115
```

```
SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Alignment Scores:
Pred. No.: 2,34e-80 Length: 1107
Score: 868.00 Matches: 178
Percent Similarity: 67.68% Conservative: 67
Best Local Similarity: 49.17% Mismatches: 115
Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2

US-09-921-992-50 (1-372) x ABQ67966 (1-1107)

QY 10 ArgLysSerThrArg---lIeTyrValGlyAsnValProIleGlyAspGlyAlaProfile 28
DB 19 CGCGAAACACCTCGCCCGCTCCCAAGTGGGTAAATTAACCTATTGGTGTAGTGGGAATTA 78
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
DB 79 ACTATCCAAAGCATGACTACTACTAAGACACACATGATGTCGAAGCAACAGTAGCAGAAAT 138
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValAtgValSerValProThrMetAspAla 68
DB 139 CACCGATTAGAGAAGCTGGTGTGAGATGGCGAGTTGCTTGCTCCTGATGAACGTGCA 198
QY 69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
DB 199 GCGAATGCTCTTAGTGCCATCAAGAAAAAGATCATATTCCGCTTGGCCACATATTCA 258
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 259 TTTGATTACCGACTAGCACTTAAGACTATTGATGCAGGTTGTGACAAAAATTCGATTAT 318
QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 319 CTGTGTAACTTTGGTGGCGCTGATCGGTGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 378
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
DB 379 AATATTCCAATCGTATTGGGTAAATGCTGTGTAGTTAGAAAAAGAAATATTCAAAAA 438
QY 149 TyrGlyGluProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 439 TATGTTACTCCCTACTGCTGACGGAATGTGAGAAAGTGCACTTGCCCATATTAAAAATCTC 498
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 499 GAAGATTATGATATCATTTGTTCTTTTGAAGGCTTCTGATGTGAATTTAGCA 558
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 559 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATCTCTGCATCTCGGAATTACA 618
QY 209 GluAlaGlyAlaAArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
DB 619 GAATCTGTGTACAAATTTGCTCGAGGAATAAAAGTGCTGCTGTTTGGAGCGGATACTC 678
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAAspProValGluGluLe 248
DB 679 AGTTTGGCGATTGGAAATACATTACAGATATCTTTGAGTGCTGATGCTGCTGGAGAAATA 738
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 739 AAAGTGGCCGGAGAGTTTAAATCAATTTGCTCTTCTCGAATCGCGCATGCTTATC 798
QY 269 AlaCysProThrCysSerArgGlnGlnPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 799 TCCTGCCCTACTTGGGTGCAATAGATATTGATTAAATTCGTATCGCTAATGAAGTGAA 858
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 859 AATTACATGACCAAGATTGAAGTTCGATTAAGTAGCCGTGCTGGTGTGCGGTCAAC 918
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328

1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Db 919 GGCCCTGGAGAAGCTCGCGAAGCCGATATCGGAATTCGTGTTCAACGAGAGGCCTT 978
QY 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAspMetIleAspGlnLeu 348
DB 979 CTTTTTAGACATCGGTAAATATTTCGAAAGAGTACCGGAAGCTATTATGATACGAACTT 1038
QY 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgIleAspValGln 368
DB 1039 AAGAAAGAAATTTGATATTTTGGCAGAGAATTT---TTTGTGAAGAAATAGATTGGAA 1095

369 GlnVal 370
1096 AGCCTT 1101

RESULT 9
ABQ69993
ID ABQ69993 standard; DNA; 1107 BP.
XX ABQ69993;
AC ABQ69993;
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes EGDe DNA sequence #205.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
infection; ds.
XX
OS Listeria monocytogenes EGDe.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
WPI; 2002-332479/37.
XX
New genomic sequences from Listeria species, useful for detection,
treatment and prevention of infection, also related polypeptides,
antibodies and modulators
Claim 16; SEQ ID 2806; 180pp; French.
The present invention relates to nucleic acid sequences
(ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
and primers for identification and/or detection of Listeria (e.g. as
contaminants in foods, or mutational analysis) and for analysis of
gene expression. Proteins encoded by the nucleic acid sequences can be
used to screen for compounds that modulate gene expression, replication
and pathogenicity of Listeria (potential therapeutic agents), also for
treating infections by Listeria, and are useful as immunogens in
anti-Listeria vaccines.
Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Alignment Scores:
Pred. No.: 2,34e-80 Length: 1107
Score: 868.00 Matches: 178
Percent Similarity: 67.68% Conservative: 67
Best Local Similarity: 49.17% Mismatches: 115
Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2
```

CC AAF28514-AAF28554). The library has a number of uses described in the  
 CC specification e.g. is useful for identifying diagnostic and therapeutic  
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large  
 CC aerobic, gram-negative diplococcus, normally found among the bacterial  
 CC flora of human upper airways. M. catarrhalis is known to cause acute,  
 CC localised infections such as otitis media, sinusitis and bronchopulmonary  
 CC infection and life-threatening, systemic diseases including endocarditis  
 CC and meningitis.

XX SQ Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;

#### Alignment Scores:

Pred. No.: 2,81e-109 Length: 33140  
 Score: 1166.00 Matches: 220  
 Percent Similarity: 78.92% Conservative: 72  
 Best Local Similarity: 59.46% Mismatches: 76  
 Query Match: 62.49% Indels: 2  
 DB: 22 Gaps: 1

US-09-921-992-50 (1-372) x AAF28536 (1-33140)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
 DB 21741 ATACAAT-----CGATTAAACCGCCCTTACCACAAAAAATCATGTGGCAATGTC 21794  
 QY 21 ProfileGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
 DB 21795 GCCATCGCGCGGATGACCCATCGCGTGCAGATATGACAAATACAAACACCTGTGAT 21854  
 QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
 DB 21855 ATTGATGACACAGTTGCACAAATTCAGCGATGCGTGCAGCGACGCGACTGTGATGCGT 21914  
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80  
 DB 21915 GTATGACACCAACGATGAGAGCGTGCCTTTGCTGAGATAAAAAACCGGTATCC 21974  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
 DB 21975 ATTCCACTCATTCGCCATATTCATTGATCACAANAATGCCATTCGATGCGCGATGTC 22034  
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120  
 DB 22035 GGTGACAGACTCGCTGATCAATCAATCAGGTAAATATTGGCAATGATCAAAAAGTCAAGAA 22094  
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
 DB 22095 GTGGTTGACGCGCAGCCATCATATAATGTCGCCATTCGATCGGTGTAATGCAAGCTCA 22154  
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyProThrProGlnAlaLeuLeuGluSer 160  
 DB 22155 CTTGAAAAGGATTTACAAAAAATAATCGAGAACCGACAGCGTGCAGCGATGCTAGATCG 22214  
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
 DB 22215 GCGTTACGCGCACATGATATTTAGAAAATCTTAATTTTCAGAAATATAAAATCTCAGTC 22274  
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
 DB 22275 AAAGCCTCAATGATTTTAAACCCCTTGATGCGTATCGATGATTTACAGCAAAATGAT 22334  
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220  
 DB 22335 AATCCATTCGATCTTGGCTTACCGAAGCGGGGTATATCGCATCGCGCGTCAATCT 22394  
 QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
 DB 22395 GCCATTGCTTTGGCGGCTTTGCTACTGTGATGCGATGGCATACCATCGTATCTCTTTG 22454  
 QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLysSerIleuArgIle 260  
 DB 22455 GCGGACACGACCAAGAAAGAGATTAATAATTTGGCTTTGACATTTAAATAATCACTGGGTATT 22514

QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
 DB 22515 CGTAGTAATGGCGTCAATTTTATTCTTGTCCAGTTGTAGCCGTCAAGAGTTTGATGTG 22574  
 QY 281 IleGlyThrValAsnAlaLeuGlnArgLeuGluAspIleIleThrProMetAspVal 300  
 DB 22575 ATCAAAAGTGAATGAGCTTGAAGCTCGTTAGAGGATATTCTGTGAGCCACTAGATCTG 22634  
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
 DB 22635 TCAGTTATTTGGCTGTAAAGTCAATGGCCAGGTGAAGCCAAAGACGACGATCGGTGTC 22694  
 QY 321 ThrGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
 DB 22695 GTGGGAGCCGCCCAATAGTTTGGTTTATTAACATGTTGTCAAAAGCCATCTGATTGAT 22754  
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 DB 22755 ACCAAAAATTTGGTAGATGAGATTGAGCGGTGGTGGTCTCAAGTCAAGACATTCAA 22814  
 QY 361 GluAlaArgArgIleAspValGlnGlnVal 370  
 DB 22815 GATAACGGCGCAATGAGATTATCCGTATA 22844  
 RESULT 8  
 ABQ67966  
 ID ABQ67966 standard; DNA; 1107 BP.  
 XX AC ABQ67966;  
 XX DT 29-AUG-2002 (first entry)  
 XX DE Listeria monocytogenes EGD DNA sequence #90.  
 XX KW Antibacterial; Listeria; food contamination; mutational analysis;  
 XX KW infection; ds.  
 XX OS Listeria monocytogenes EGD.  
 XX PN WO200228891-A2.  
 XX PD 11-APR-2002.  
 XX PF 04-OCT-2001; 2001WO-PR03061.  
 XX PR 04-OCT-2000; 2000FR-0012697.  
 XX PA (INSP ) INST PASTEUR.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Kunst F, Glaser P;  
 XX WPI; 2002-332479/37.  
 XX DR  
 XX PT New genomic sequences from Listeria species, useful for detection,  
 XX PT treatment and prevention of infection, also related polypeptides,  
 XX PT antibodies and modulators  
 XX PS Claim 16; SEQ ID 779; 180pp; French.  
 XX CC The present invention relates to nucleic acid sequences  
 XX CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes  
 XX CC and primers for identification and/or detection of Listeria (e.g. as  
 XX CC contaminants in foods, or mutational analysis) and for analysis of  
 XX CC gene expression. Proteins encoded by the nucleic acid sequences can be  
 XX CC used to screen for compounds that modulate gene expression, replication  
 XX CC and pathogenicity of Listeria (potential therapeutic agents), also for  
 XX CC treating infections by Listeria, and are useful as immunogens in  
 XX CC anti-Listeria vaccines.  
 XX CC Note: The sequence data for this patent did not form part  
 XX CC of the printed specification, but was obtained in electronic format  
 XX CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.



PR 07-JUN-1995; 95US-0487429.  
 PR 21-APR-1995; 95US-0426787.  
 PR 07-JUN-1995; 95US-0476102.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (UJO) UNIV JOHNS HOPKINS.  
 XX  
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;  
 XX WPI; 1996-485782/48.  
 DR  
 XX Haemophilus influenzae Rd genome recorded on computer readable  
 PT medium - useful for identifying commercially important nucleic acid  
 PT fragments by homology searching  
 XX  
 PS Claim 1; Page 77.2-77.1091; 1291pp; English.  
 XX  
 CC This sequence represents the complete genome sequence of the bacterium  
 CC Haemophilus influenzae strain Rd. The invention relates to a computer  
 CC readable medium (CRM) having recorded upon it the complete H. influenzae  
 CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide  
 CC sequence at least 99% identical to (I). By providing the full-length  
 CC genomic sequence in a computer readable form, it is possible to identify  
 CC commercially important nucleic acid fragments and expression modulating  
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to  
 CC regulate the expression of a nucleic acid molecule. Vectors and altered  
 CC organisms comprising the predicted ORFs can be used to produce any of the  
 CC polypeptide fragments of the H. influenzae Rd genome.  
 XX  
 SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:  
 Pred. No.: 2,6e-152 Length: 1830121  
 Score: 1603.50 Matches: 320  
 Percent Similarity: 93.66% Conservative: 20  
 Best Local Similarity: 88.15% Mismatches: 22  
 Query Match: 85.93% Indels: 1  
 DB: 17 Gaps: 1

US-09-921-992-50 (1-372) x AAT42063 (1-1830121)

QY 4 GlnAlaProLeuGlnArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23  
 DB 390972 CAGCCCAACTATTAAAGCGTCGTGAATCGACAAAAATTTATGTGGAAATGTACCAATTTGGT 391031  
 QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43  
 DB 391032 GGGGATGCGCCTATTGCGGTGCAATCAATGACAAATCTCGCACCACTGATGTGGAGCG 391091  
 QY 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63  
 DB 391092 ACAGTTGCTCAATTAATCAATTAGAACGTTGTTGGTCAGATATTGTTGCTGTATCTGTT 391151  
 QY 64 ProThrMetAspAlaGluAlaPheLysLeuIleLysGlnValAsnValProLeu 83  
 DB 391152 CCACCAATGGATGTCGGGAAGCATTTAAACAAATTAACAAAGTGAATGTTCCGCTC 391211  
 QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103  
 DB 391212 GTAGCAGATATTCAATTCGACTATCGTATCGGTAAAGTCGACAGATATGAGTGGAT 391271  
 QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123  
 DB 391272 TGTTTACGTATCAATCTCGCAACATTTGTCGTGAAGATCGCGTGGCTGTTGAT 391331  
 QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143  
 DB 391332 TGTGCGCGACACAAAATATTCGATTTCGATTGGTGAATGCAAGCTCTTTAGAAAAA 391391  
 QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSerAlaMetArg 163  
 DB 391392 GATTTCAGAAAAAATATGCGCAACCAACCGCCAGACCTTGTAGAAATCCGATTCGCGT 391451

QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183  
 DB 391452 CATGTAGAAATTCAGATCGTCTTAACATCGATCAGTTAAAGTCGAGCGTAAAGCGCTCC 391511  
 QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203  
 DB 391512 GATGTATTCTTAGCGGTGAATCTTATCGTTTACTGGCTAAAGCAATTAACAGCGCTTTA 391571  
 QY 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyValValLysSerAlaIleGly 223  
 DB 391572 CATTTAGGCATTACAGAGCAGGTGGCGCACCGCTGGTGCAGTAAATCTCGCATGGGT 391631  
 QY 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAsp 243  
 DB 391632 TTAGGAATGTTATTAGCTGAGGCGCATTTGGCGATACACTACGCGTCTCTTTGGCGGAGAT 391691  
 QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263  
 DB 391692 CCTGTAGAGGAATCAAAAGTCGGTTTGATATTTTGAATCTTTACGGATTCGTTCAAGA 391751  
 QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283  
 DB 391752 GGAATTAACTTTATTTGCTTGCCCAACCTGTTCTGCCAAGAATTTGATGTAATCGGTACA 391811  
 QY 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303  
 DB 391812 GTAAATGCGCTAGAACACACGCTTGAAGATATTTATTACCAATGATGATCTATTATC 391871  
 QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323  
 DB 391872 GGTGTGTAGTGAATGGTCTCGCGAGGCACCTCGTCTCGGATCTCGGCGTAACGGCGGT 391931  
 QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsn 342  
 DB 391932 AACAAAAAAGCGTTATTATCTTGACGGAGAACGCCAAAAGAGCGTTTGTATACGAA 391991  
 QY 343 AspMetIleAspGlnLeuGluAlaArgIleAlaLysAlaSerGlnLeuAspGluAla 362  
 DB 391992 GATATAGTGAACCAATTTAGAACAAAAATTCGTGCGAAAGTCGCACGACAGATCCAAAA 392051  
 QY 363 ArgArgIle 365  
 DB 392052 AACAGAAAT 392060  
 RESULT 6  
 ID ABA92787 standard; DNA; 640681 BP.  
 AC ABA92787;  
 DT 27-MAR-2002 (first entry)  
 DE Buchnera sp. genomic DNA SEQ ID NO:1.  
 KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;  
 KW circular; ds.  
 OS Buchnera sp.  
 PN JP2001292771-A.  
 XX 23-OCT-2001.  
 PD 07-APR-2000; 2000JP-0107160.  
 PF 07-APR-2000; 2000JP-0107160.  
 PR (RIKA) RIKAGAKU KENKYUSHO.  
 PA WPI; 2002-126043/17.  
 DR A genomic DNA of cockroach-symbiotic bacterium -  
 XX  
 XX

XX PA (MONS ) MONSANTO TECHNOLOGY LLC.

XX PI Boronat A., Campos N., Rodriguez-concepcion M., Rohmer M., Seeman M;

XX PI Valentin HE, Venkatesh TV, Venkatramesh M;

XX DR P-PSDB; AAEE19653.

XX DR WPI; 2002-227151/28.

XX PT gcpE nucleic acid which is an essential gene of the methyl-D-erythritol

XX PT phosphate pathway, encoding a fully defined GCPE protein which is

XX PT useful for increasing levels of tocopherol substrates in plants

XX PS Claim 3; Page 117-119; 155pp; English.

XX CC The invention relates to gcpE nucleic acid molecule, an essential gene

XX CC of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice,

XX CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful

XX CC for producing a transgenic plant such as Brassica campestris, B. napus,

XX CC canola, castor bean, coconut, cotton, crame, linseed, maize, mustard,

XX CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,

XX CC or wheat with an increased isoprenoid (tocopherol) compound level. The

XX CC expression of GCPE protein in organisms increases the level of

XX CC tocopherol substrate such as isopentyl diphosphate and dimethylallyl

XX CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE

XX CC protein can nutritionally enhance food and feed sources. Overexpression

XX CC of GCPE protein in transgenic plant may provide tolerance to stresses

XX CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV

XX CC tolerance, etc. GCPE may be used to obtain nucleic acid molecules from

XX CC the same species, and to obtain nucleic acid homologues. gcpE is also

XX CC used as or primers. The recombinant vectors are used in plant

XX CC transformation or transfection. gcpE an also act as markers capable of

XX CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).

XX CC gcpE is also used to determine the level or pattern of expression of

XX CC the protein. The present sequence is Escherichia coli gcpE gene.

XX SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:

Pred. No.:	7.12e-184	Length:	1119
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-09-921-992-50 (1-372) x AAD31203 (1-1119)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20

DB 1 ATGCATAACGAGCTCCATTCCACGTAGANAATCAACAGCTATTACCTGGGAATGTG 60

QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40

DB 61 CCGATTGGCGATGGTCTCCATCGCGTACAGTCCATGACCAATACGGTACGACAGAC. 120

QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60

DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180

QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80

DB 181 GTATCCGTACCGAGTATGACGGCGGAGAGCGTTCAACTCATCAACAGCAGGTTTAC 240

QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100

DB 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGGAATAC 300

QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluAlaArgIleArgMet 120

DB 301 GCGCTCGATTGTCTGCGTATTACCCCTGGCAATATCGGTAATCAAGAGCGTATTCGCATG 360

QY 121 ValValAspCysAlaAlaAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140

DB 361 GTGGTTGACTGTGCGCGCATAAAAACATTCCGATCCGTTATTTGGCGTTAAACGCCGATCG 420

QY 141 IeuGluLysAspLeuGlnIleLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

DB 421 CTGGAAAAAGATCTCGAAGAAAAGTATGCGAACCGACCGCCGAGCGGTTGCTGGAAATCT 480

QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180

DB 481 GCATGCGTCACTGTTGATCATCTCCATCGCCTGAACCTTCGATCAGTTCAAGTCAGCGTG 540

QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200

DB 541 AAAGCGTCTGACGTCTTCTCGCTGTTGAGTCTTATCGTTCTGGCAAAACAGATCGAT 600

QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220

DB 601 CAGCCGTTGTCATCTGGGATCAACGAACCGGTGGTGGCGGAGCGGGCAGTAAATCC 660

QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

DB 661 GCATTTGGTTTAGGTCCTGCTGCTGTAAGGCATCGGCGACACCGCTGCCGTATCGCTG 720

QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260

DB 721 CGCGCGGATCCGTCGAAGAGATCAAGTCGGTTCGATATTTTGAATTCGCTGCGTATC 780

QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280

DB 781 CGTTCCGAGGAGATCAACTTTCATCGCCTGCCGACCTGTTCCGCTCAGGAATTTGATGTT 840

QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300

DB 841 ATCGGTAGCTTTAAACGCTGGAGCAACGCTTGAAGATATCATCTCCGATGACGTT 900

QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320

DB 901 TCGATTATCGGCTGCTGTTGTAATGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960

QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340

DB 961 ACCGGCGGCACACAGAAAAGCGCCTCTATGAAGATGCGCTGCGCAAGACCGTCTGGAC 1020

QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360

DB 1021 AACAACGATATGATCGACCGCTGGAAGCACCATTCGTGCGAAGCCAGTCAGCTGGAC 1080

QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372

DB 1081 GAAGCGCTCGAATTGACGTTACAGCAGGTGAAAAA 1116

RESULT 5

AAT42063

ID AAT42063 standard; DNA; 1830121 BP.

XX AC AAT42063;

XX DT 14-SEP-1999 (first entry)

XX DE Haemophilus influenzae complete genome sequence.

XX KW Genome; bacterium; Haemophilus influenzae; computer readable medium;

XX KW expression modulating fragment; regulation; gene expression; vector;

XX OS Haemophilus influenzae.

XX PN WO9633276-A1.

XX PD 24-OCT-1996.

XX PF 22-APR-1996; 96WO-US05320.



XX WPI: 2001-025196/03.  
 DR P-PSDB; AAB45692.  
 XX  
 PT Incorporating gcpE and yfgB genes into viruses and cells, for  
 PT increasing isoprenoid content and identifying e.g. antimicrobial  
 PT agents, comprises using DNA sequences from bacteria or parasites  
 XX  
 PS Claim 3; Page 13-15; 36pp; German.  
 XX  
 CC This invention describes a novel method for incorporating gcpE and yfgB  
 CC genes into viruses and cells for increasing isoprenoid content and  
 CC identifying e.g. antimicrobial agents, comprises using DNA sequences (I)  
 CC from the gcpE or yfgB genes of bacteria or parasites or DNA sequences (II)  
 CC which hybridize to the specified genes or encode a plastid protein  
 CC with the same biological activity as those encoded by the genes. The  
 CC invention also describes (1) plant cells containing (I) or (II); (2)  
 CC transformed plant cells, and transgenic plants regenerated from them,  
 CC that contain (I) or (II); (3) determining the enzymatic activity of a  
 CC gcpE protein; or (4) screening compounds (A) that have antimycotic,  
 CC antiparasitic or antiviral activity in humans or animals or antiviral,  
 CC antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)  
 CC are used: (i) to increase the isoprenoid levels in viruses and cells;  
 CC (ii) for determining the enzymatic activity of gcpE and yfgB proteins;  
 CC and (iii) to identify compounds that inhibit activity of gcpE, i.e.  
 CC potential antibacterial, antimycotic, antiparasitic or antiviral agents  
 CC for use in humans or animals, or antiviral, antiparasitic, fungicidal  
 CC or herbicidal agents for agriculture.  
 XX  
 SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
 XX

Alignment Scores:  
 Pred. No.: 7,12e-184 Length: 1119  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-921-992-50 (1-372) x AAC82653 (1-1119)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
 DB 1 ATGCATAACCCAGGCTCCAAATTCACGTAGAAATCAACAGTATTACGTGGGAATGTG 60  
 QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40  
 DB 61 CCGATTGGCGATGGTGTCCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC 120  
 QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
 DB 121 GTCGAAGCAACGGTCAATCAATCAAGCGCTGGAACGGCTTGGCGTGTATATCGTCCGT 180  
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnInValAsn 80  
 DB 181 GTATCCGTACCGAGTGGACGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAAC 240  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
 DB 241 GTGCCCGTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGGAATAC 300  
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
 DB 301 GGGCTGCAATGTCTGGGTATTAACCTTGGCAATATCGTAAATGAAGCGGTATTCGCATG 360  
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
 DB 361 GTGGTTGACTGTGGCGCGATNAACATTCAGATCCGATCCGATTGGCGTTAACGCCGATCG 420  
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
 DB 421 CTGGAAGAAAGATCTGCAAGAAAAAGTATGGCAACCCAGCCGCGAGCGTGTCTGGAATCT 480

QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
 DB 481 GCATGGCTCATGTTGATCATCTCGATCGCTGAATTCGATCAGTTCAAAGTCAGCGTG 540  
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
 DB 541 AAAGCGCTCTGACGCTCTTCCTCGCTGTAGTCTTATCGTTCTGCGCAAAACAGATCGAT 600  
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220  
 DB 601 CAGCCGTTGTCATCTGGGATACACCAAGCCGCTGGTGGCGCAGCGGCGAGTAAATCC 660  
 QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
 DB 661 GCATTTGGTTTAGGCTCTGCTGCTGCTGAAGCATCGGCGACACGCTGCGCTATCGCTG 720  
 QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
 DB 721 GCGGCCGATCCGCTCGAAGAGATCAAAAGTCGGTTCGATATTTGAAATCGCTGCGTATC 780  
 QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
 DB 781 CATTTCGAGGATCAACTTCATCGCTGCCGACCTGTCGCGTCAGGAATTTGATGTT 840  
 QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
 DB 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGACGTT 900  
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
 DB 901 TCGATTATCGGCTCGTGTGAATGCCAGGTGAGCGCTGCTTCTACACTCGCGCTC 960  
 QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
 DB 961 ACCGCGGCAACAAAGAAAGCGCTCTATGAAGATGGCGTGGCAAAACCGCTCTGGAC 1020  
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 DB 1021 AACACGATATGATCGACCGCTGGAGCACCGATTCGTCGAAGCCAGTCACTGGAC 1080  
 QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
 DB 1081 GAAGCGCGTGAATTCAGCTTCAGCAGGTGAAAAA 1116

RESULT 4  
 AAD31203  
 ID AAD31203 standard; DNA; 1119 BP.  
 XX  
 AC AAD31203;  
 DX 31-MAY-2002 (first entry)  
 XX  
 DE Escherichia coli gcpE gene.  
 XX gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;  
 KW transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;  
 KW food; feed source; transfection; single nucleotide polymorphism; SNP;  
 KW oxidative stress tolerance; UV tolerance; transformation; plant; ds.  
 XX Escherichia coli.  
 OS  
 FH Key Location/Qualifiers  
 FT CDS 1..1119  
 FT /\*tag= a  
 FT /product= "Escherichia coli GCPE protein"  
 XX  
 XX WO200212478-A2.  
 XX  
 XX 14-FEB-2002.  
 PD  
 XX  
 XX 06-AUG-2001; 2001WO-US24335.  
 PF  
 XX  
 XX 07-AUG-2000; 2000US-223483P.  
 PR

DR WPI; 2000-640125/62.  
XX  
PT Identifying antagonists of the expression of gene encoding bacterial  
PT growth polypeptide useful for treating bacterial infections or  
PT diseases, by evaluating transcription of the gene in the presence of  
PT test molecule -  
XX  
PS Claim 1; Page 25; 55pp; English.  
XX  
CC The present sequence is that of the Escherichia coli FUN gene  
CC gcpE, which encodes a protein that is essential for bacterial  
CC growth or survival. gcpE is 1 of 22 E. coli genes (see  
CC AAA8692-713) identified as being essential (there is no deletion  
CC genotype). These 22 genes fulfil criteria for being attractive  
CC antibacterial targets: hypothetical open reading frames coding for  
CC essential functions (mutation is lethal for growth in rich media);  
CC broad conservation (orthologues are present in a wide range of  
CC bacteria, including Haemophilus influenzae, Streptococcus pneumoniae,  
CC Helicobacter pylori and Borrelia burgdorferi); and low toxicity  
CC potential in higher organisms (mostly no orthologues were identified  
CC in Saccharomyces cerevisiae). An antagonist or inhibitor of the  
CC expression of an essential gene or of its function provides the key  
CC for antibacterial therapy. The invention provides methods for  
CC identifying such antagonists or inhibitors. These involve  
CC contacting a bacterial cell comprising an essential gene with a  
CC candidate antagonist or inhibitor, and testing whether contact leads  
CC to cell growth inhibition and/or cell death. The method allows the  
CC development of new broad spectrum antibiotics. A conditional mutant  
CC of an essential gene can be used to induce a lethal phenotype in  
CC bacteria for the analysis of surrogate markers.  
XX  
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
  
Alignment Scores:  
Pred. No.: 7,12e-184 Length: 1119  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0  
  
US-09-921-992-50 (1-372) x AAA88705 (1-1119)  
  
QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
DB 1 ATGCATACCAGGCTCCATTCACCTAGAGTAAGAAATCAACAGCTATTACGTTGGGAATGG 60  
  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40  
DB 61 CGGATTGGCGATGGTCTCCCATCGCGTACAGTCCATGACCAATACGGGTACGACAGAC 120  
  
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
DB 121 GTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180  
  
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80  
DB 181 GATTCGTTACCGACGATGACGGCGGAGAGCGGTTCAAACTCATCAACAGCAGGTTAAC 240  
  
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
DB 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCGCTGAAGAGTACGGGAATAC 300  
  
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120  
DB 301 GCGCTCGATTGTCTGCGATTATACCTCGCAATATCGTAATGAAGAGCGTATTTCGATG 360  
  
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
DB 361 GTGGTTGATCTGCGCGCGGATATAAAACATTCGATCCGATTGGCGTTAACGCCGATCG 420  
  
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
|||

DB 421 CTGGAAGAAAGATCTGCAAGAAAAGTATGCGCAACCGACCGCGCGCTTGTGGAATCT 480  
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
DB 481 GCATGCGTCAATGTTGATCATCTCGATCCCTGAACTTCGATCAGTTCAGAAATCAGCGTG 540  
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
DB 541 AAAGCGCTGACGCTCTCTCTCGCTGTTGAGTCTTATCGTTTGTGCAACACAGATCGAT 600  
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220  
DB 601 CAGCCCTTCATCTGGGATCACCCGAAGCGGTGGTGGCGCGAGCGGCGAGTAAATCC 660  
QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
DB 661 GCATTTGGTTTGGTCTGCTGCTGTAAGGATCGCGGACACGCTGCGCGTATCGCTG 720  
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
DB 721 GCGGCGATCCCGTCCGAGAGATCAAGTCGGTTTCGATATTTGAAATCGCTGCGTATC 780  
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
DB 781 CGTTCCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAATTTGATGTT 840  
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
DB 841 ATCGGTACGGTTAACGCGCTGGAGCAACGCCCTGGAAGATATCATCACTCCGATGGAAGT 900  
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
DB 901 TCGATTATCGGCTGGCTGGTGAATGGCCAGGTGAGCGCTGTTTCTACACTCGGCGTC 960  
QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
DB 961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGCTCTGGAC 1020  
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
DB 1021 AACAAAGATATGATCGACCGCTGGAAGCAGCATTCGTCGGAAGCCAGTCAGCTGGAC 1080  
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
DB 1081 GAAGCGCTCGAATTGACGTTGACGAGGTGGAATAA 1116  
  
RESULT 3  
AAC82653  
ID AAC82653 standard; DNA; 1119 BP.  
XX  
AC AAC82653;  
XX  
DT 15-MAR-2001 (first entry)  
XX  
DE E. coli gcpE DNA.  
XX  
KW Isoprenoid; gcpE; YfgB; antimicrobial; transgenic plant; agriculture;  
KW antimycotic; antiparasitic; antiviral; fungicidal; herbicidal; ds.  
XX  
OS Escherichia coli.  
XX  
PN WO200072022-A1.  
XX  
PD 30-NOV-2000.  
XX  
PF 20-MAY-2000; 2000WO-EP04592.  
XX  
PR 21-MAY-1999; 99DE-1023567.  
PR 21-MAY-1999; 99DE-1023568.  
XX  
PA (JOMA/) JOMAA H.  
XX  
PI Jomaa H;

XX WO20061793-A2.  
PN 19-OCT-2000.  
XX 07-APR-2000; 2000WO-EP03135.  
XX 09-APR-1999; 99EP-0107031.  
PR 04-FEB-2000; 2000EP-0102111.  
XX (GPCB-) GPC BIOTECH AG.  
PA Loferer H, Jacobi A;  
XX WPI; 2000-697048/67.  
DR  
XX  
XX Identifying antibacterial compounds, comprises identifying an  
PT antagonist or inhibitor of the expression of a gene encoding a  
PT polypeptide essential for bacterial growth or survival -  
XX  
XX Claim 1; Fig 1; 75pp; English.  
XX  
XX The present invention relates to antagonists and inhibitors of 24  
CC bacterial genes and proteins. The proteins are thought to be essential  
CC for growth in several species of bacteria (including *S. pneumoniae*, *B.*  
CC *burgdorferi*, *H. influenza* and *H. pylori*). The proteins and coding  
CC sequences shown in the specification can be used to identify antagonists  
CC and inhibitors which can be used in disease treatment and pesticides. In  
CC particular, they can be used against *M. tuberculosis*. The present  
CC sequence is one of the genes of the invention.  
XX  
XX Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
SQ  
Alignment Scores:  
Pred. No.: 7,12e-184 Length: 1119  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0  
US-09-921-992-50 (1-372) x AAA95478 (1-1119)  
QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
DB 1 ATGCATAACGAGCTCCAAATCAACGTAGAAATCAACACGTAATTCAGTTGGGAATG 60  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
DB 61 CCGATTGGCGATGGTCTCCATCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120  
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
DB 121 GTCGAAGCAACGGTCAATCAAAATCAAGCGCTGGAACCGCTGGCGCTGATATCGTCCGT 180  
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheIleLeuLysGlnGlnValAsn 80  
DB 181 GTATCCGTACCGACGATGGACGGCGGAGACGCTTCAAACTCATCAACAGCGGTAAAC 240  
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
DB 241 GTCCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCGTGAAGTAGCGGATAC 300  
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
DB 301 GGCGTGGATTGTCTGGCTATTAAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTCGCATG 360  
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
DB 361 GTGGTTGACTGTGGCGCGATAAACATTCCTCGATCCGATTCGCGTTAACCGCGGATCG 420  
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

DB 421 CTGGAAAAAGATCTGCAAGAAAAAGTATGGCGAACCGCCGCGCGCTTGTGGAATCT 480  
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
DB 481 GCCATGCGGTCAATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540  
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
DB 541 AAAGCGTCTGAGCTCTTCTCGCTGTTGAGTCTTATCGTTGTCGCAAAACAGATCGAT 600  
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220  
DB 601 CAGCGCTTCATCTGGGATCACCGAAGCGGTGGCGCGCGGCGGCGAGTAAATCC 660  
QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
DB 661 GCCATTGGTTTGGTCTGCTGCTGAAGGCATCGGCGACACGCTGGCGGTATCGCTG 720  
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
DB 721 GCGGCGGATCCGTCGAAGAGATCAAGTCGTTTCGATATTTTGAATCGCTGCGTATC 780  
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
DB 781 CGTTCCGCGAGGATCAACTTCATCGCTGCCCGACCTGTTCGCGTCAGGAATTTGATGTT 840  
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
DB 841 ATCGGTACGGTTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCCTCCGATGGACGT 900  
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
DB 901 TCGATTATCGGCTGGTGGTGAATGCCCGAGGTGAGCGCTGTTTCTACACTCGCGCTC 960  
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340  
DB 961 ACCGGCGGCAACAAGAAAGCGCGCTCTATGAAGATGGCGTCGCGAAAGACCGTCTGGAC 1020  
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
DB 1021 AACAACGATATATGACCGCTGGAAGCACCGATTCGTGGCAAGCCAGTCAGCTGGAC 1080  
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
DB 1081 GAAGCGCTCGAATTGACGTTTCAGCGTTGAAAAA 1116  
RESULT 2  
AAA88705  
ID AAA88705 standard; DNA; 1119 BP.  
XX  
AC AAA88705;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE E. coli FUN essential gene gcpe.  
XX  
KW FUN gene; gcpe gene; essential gene; antibacterial; antibiotic;  
XX screening; infection; therapy; antagonist; surrogate marker; ds.  
OS Escherichia coli.  
XX  
XX EP1043403-A1.  
XX  
PD 11-OCT-2000.  
XX  
XX 09-APR-1999; 99EP-0107031.  
XX  
XX 09-APR-1999; 99EP-0107031.  
XX (GPCB-) GPC GENOME PHARM CORP AG.  
XX  
XX Not given;  
XX

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: November 23, 2003, 17:17:14 ; Search time 336 Seconds  
(without alignments)  
2988.664 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MNQAPIQRKSTRIYGVN.....RAKASQLDEARRIDVQVEK 372

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_p2n.model -DRV=xlh  
-Q=/cg2\_1/USPTO.spool/US09921992/runat\_21112003\_162545\_18397/app\_query.fasta\_1.519  
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NOEM-EXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09921992.@CGN\_1\_0 @runat\_21112003\_162545\_18397 -NCPUS=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N Geneseq 19Jun03.+:  
1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1866	100.0	1119	21	AAA95478	E. coli essential
2	1866	100.0	1119	21	AA888705	E. coli FUN essent
3	1866	100.0	1119	22	AA82653	E. coli gcpE DNA.
4	1866	100.0	1119	24	AAD31203	Escherichia coli g
5	1603.5	85.9	1830121	17	AAT42063	Haemophilus influe
6	1282	68.7	640681	24	ABA92787	Buchnera sp. genom
7	1166	62.5	33140	22	AAF28536	Genomic fragment #
8	868	46.5	1107	24	ABQ67966	Listeria monocytog
9	868	46.5	1107	24	ABQ65993	Listeria monocytog
10	868	46.5	2944528	24	ABA03041	Listeria monocytog
11	861.5	46.2	6157	24	ABQ70939	Listeria monocytog
12	822	44.1	1083	24	ABK74401	Bacillus lichenifo
13	814.5	43.6	349980	24	ABQ81842	Bifidobacterium lo
14	812	43.5	994	24	ABQ68324	Listeria monocytog
15	794	42.6	1180	19	AAK14017	H. pylori GHPO 16
16	785.5	42.1	35829	23	AAS59573	Propionibacterium
17	761	40.8	1134	22	AAH67170	C glutamicum codin
18	761	40.8	1179	25	ACA00649	C. glutamicum deri
19	761	40.8	349980	22	AAH68530	C glutamicum codin
20	761	40.8	349980	22	AAH68531	C glutamicum codin
21	751.5	40.3	1161	25	ABZ71126	Mycobacterium tube
22	751.5	40.3	4403765	22	AAI99683	Mycobacterium tube
23	751.5	40.3	4411529	22	AAI99682	Mycobacterium tube
24	751.5	40.3	4411529	22	AAI99682	Mycobacterium tube
25	682	36.5	5484	23	AA573006	DNA encoding novel
26	534	28.6	1263	25	ABZ33756	N. gonorrhoeae nuc
27	527	28.2	60873	21	AA881469	N. meningitidis pa
28	527	28.2	349980	21	AAF21610	Neisseria meningit
29	527	28.2	1437668	21	AAA81490	N. meningitidis B
30	510	27.3	1857	23	AA590066	DNA encoding novel
31	490	26.3	1038602	20	AAZ01425	Complete genome se
32	470.5	25.2	273254	21	AAC81914	Chlamydia pneumoni
33	470.5	25.2	1230025	20	AAK91990	Nucleotide sequenc
34	462	24.8	2520	24	AAD31201	Arabidopsis thalia
35	460	24.7	2109	22	AAC82654	P. falciparum gcpE
36	446	23.9	3147	21	AAK82996	P. falciparum gcpE
37	428.5	23.0	507	21	AAK75580	Human OREX ORF135
38	428.5	23.0	507	24	ABN78496	Human ORF1443 cDNA
39	383	20.5	435	24	ABX66784	Helicobacter pylor
40	339	18.2	670	24	AAD31221	Zea mays partial g
41	326	17.5	596	24	AAD31222	Zea mays partial g
42	321	17.2	584	24	AAD31220	Zea mays partial g
43	300	16.1	4467	10	AAK92428	Sequence encoding
44	292	15.6	33675	24	AAX31202	Oryza sativa gcpE
45	277.5	14.9	2535	20	AAK20674	Polynucleotide seq

ALIGNMENTS

RESULT 1  
AAA95478  
ID AAA95478 standard; DNA; 1119 BP.  
XX  
AC AAA95478;  
XX  
DT 27-FEB-2001 (first entry)  
XX  
DE E. coli essential gene gcpE.  
XX  
KW Bacterial growth; inhibitor; yggB; yfhC; yacE; ychB; vejD; yrfI;  
KW yggJ; yjeE; yiaO; yrdC; yhcC; ygbC; ybeY; ybeY; gcpE; kdtB; pfs; yaaJ;  
KW B1808; yeaA; yagP; B1983; yidD; yceG; yjbc; yjbc; antibacterial compound;  
KW H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;  
KW M. tuberculosis; antibiotic; ds.  
XX  
OS Escherichia coli.

**THIS PAGE BLANK (USPTO)**

Qy 206 CAGAAAGCGTTCAAACTCATCAA 227  
| | | | |  
Db 658 CTGATGCCTGCTTTGAGATCAA 679  
| | | | |

Search completed: November 23, 2003, 15:53:54  
Job time : 2756 secs

```
Db 330 CCACGGCGGCATGATGTTGGCGGACCTTCGAGAAGATCCGTGTGAACCCCGGCAACT 389
QY 335 TCGGTAATG 343
Db 390 TCGCTGACG 398

RESULT 14
AV641368
LOCUS
DEFINITION
  AV641368 Chlamydomonas reinhardtii 5% CO2 linear EST 15-DEC-2000
  cDNA clone HCL032409_r 5', mRNA sequence.
ACCESSION
  AV641368
VERSION
  AV641368.1 GI:10784696
KEYWORDS
  EST.
SOURCE
  Chlamydomonas reinhardtii
  Chlamydomonas reinhardtii
  Chlamydomonas reinhardtii
  Chlamydomonadaceae; Chlamydomonas.
REFERENCE
  1 (bases 1 to 377)
  Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohya, K.,
  Nakamura, Y., and Tabata, S.
  Generation of expressed sequence tags from low-CO2 and high-CO2
  adapted cells of Chlamydomonas reinhardtii
  DNA Res. 7 (5), 305-307 (2000)
20539644
MEDLINE
11089912
COMMENT
  Contact: Erika Asamizu
  The First Laboratory for Plant Gene Research
  Kazusa DNA Research Institute
  Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
  Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
  source
    1..377
      /organism="Chlamydomonas reinhardtii"
      /mol_type="mRNA"
      /strain="C9"
      /db_xref="taxon:3055"
      /clone="HCL032d09 r"
      /clone_lib="Chlamydomonas reinhardtii 5% CO2"
      /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
      XhoI; The cDNA library was constructed from cells cultured
      in a medium with bubbling air containing 5% carbon
      dioxide"
BASE COUNT
  89 a 118 c 116 g 54 t
ORIGIN
  1
    377
      5.5%; Score 62; DB 9; Length 377;
      Best Local Similarity 55.6%; Pred. No. 8.7e-08;
      Matches 150; Conservative 0; Mismatches 105; Indels 15; Gaps 1;
QY 50 TTGGGATGCGGATGCGATGGTGTCTCCATCGCGTACAGTCCATGACCAATACGC 109
Db 102 TCGGCAAGTGGCGGTGGGCGGAGCAGCAGCGATCGCCCTTCAGACATACCAACCGG 161
QY 110 GTACGACAGACGTGCAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGTTGGCGCTG 169
Db 162 ACACCGCATGTGCGAGTGCAGTGCAGCGACGAGTGAAGAGTGGCGGAGCGCGCGG 221
QY 170 ATATCGTCCGTGTATCGGTACCGACGATGACGCGGAGAGCGTTCA----- 217
Db 222 ACATCGTCCGCATCACAGTGCAGGCGAAGAGGAGGCTGAGGCGTGCATGAAGATTGCGG 281
QY 218 ---AATCATCAACAGCAGGTAACTGCGCGTGGTGGCTGCATCCACTTCGACTATC 274
Db 282 AGCAGTGTTCAGAGCAAGTACGAGTGCCTTCCTGTCGCGGACATCCACTTCACGCCCA 341
QY 275 GCATTGCGCTGAAAGTAGCGGAATACGGCG 304
Db 342 CGGTGGCCATGATGTTGGCGGAGCGCCTTCG 371

RESULT 15
```

```
CD424926
LOCUS
DEFINITION
  SAL_9 D05_g1 A002 Salicylic acid-treated seedlings Sorghum bicolor
  cDNA Clone SAL_9_D05_A002 5', mRNA sequence.
ACCESSION
  CD424926
VERSION
  CD424926.1 GI:31331189
KEYWORDS
  EST.
SOURCE
  Sorghum bicolor (sorghum)
  Sorghum bicolor
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
  clade; Panicoideae; Andropogoneae; Sorghum.
REFERENCE
  1 (bases 1 to 688)
  Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein
  , R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan
  , N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O.,
  Eastman, A. and Pratt, L.H.
  An EST database from Sorghum: salicylic acid-treated seedlings
  Unpublished
  Other ESTs: SAL_9 D05.bl A002
  Contact: Cordonnier-Pratt MM
  Laboratory for Genomics and Bioinformatics
  The University of Georgia, Department of Plant Biology
  Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
  Tel: 706 542 1860
  Fax: 706 583 0210
  Email: mmpratt@uga.edu
  Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
  the Human Genome Center, University of Tokyo Institute of Medical
  Science; plant material and RNA prepared at Texas A & M University;
  sequencing done in the Laboratory for Genomics and Bioinformatics,
  University of Georgia. Sequence ends have been trimmed to exclude
  vector and regions below Phred quality 16. Three-prime sequences
  are presented as their reverse complement and have been trimmed to
  exclude polyA.
  Seq primer: Sugs (CTTCTGCTCTAAAGCTGCG).
FEATURES
  source
    1..688
      /organism="Sorghum bicolor"
      /mol_type="mRNA"
      /cultivar="BTx623"
      /db_xref="taxon:4558"
      /clone="SAL_9 D05_A002"
      /lab_host="DH10B-T1 phage-resistant E. coli"
      /clone_lib="Salicylic acid-treated seedlings"
      /note="Vector: pME18S-FU3; Site 1: XhoI; Site 2: XhoI; The
      library was prepared from polyA+ RNA from seedlings grown
      in hydroponic culture. At 8 days of age, medium was
      supplemented with 1 mM salicylic acid (SA). Roots and
      shoots were harvested after 27 and 72 hr and material from
      both time points was combined prior to RNA isolation.
      Double-stranded cDNA was cloned unidirectionally into
      different DraIII sites of the pME18S-FL3 vector (5'-prime
      DraIII site is CACTGTGTG, 3'-prime DraIII site is CACATGTG
      ). XhoI excises the cDNA insert."
BASE COUNT
  174 a 172 c 205 g 137 t
ORIGIN
  1
    688
      5.2%; Score 58; DB 14; Length 688;
      Query Match
      Best Local Similarity 55.4%; Pred. No. 2e-06;
      Matches 112; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 26 GTAGAAATCAACACGTATTTACGTTGGGAATGTGCGGATGGCTGCTCCCATCG 85
Db 478 GGAGGAAACTCGAAGTGTGATGGTGGGAATGTGCCACTTGGCAGTGATCATCCATAA 537
QY 86 CGGTACAGTCCATGACCAATACGCTAGCAGACAGCTGGAAGCAACGGTCAATCAATCA 145
Db 538 GGATTCAACACCATGACGACTTCAGATACCAAGAGTGTTCGCAAAAACAGTAGAGGAGTGA 597
QY 146 AGCGCTGGAACGGCTTGGCGGTGATATCGTCCGTGTATCCGTACCGACGATGGACGCG 205
Db 598 TGAGGATACGAGATAAAGGAGCTGATATTGTTAGATAACCGTTTCAGGTGAGGAAGAG 657
```





```

156 ACACCCGCAATGTGACGTGACCGTGACCGAGTGAAGTGCAGCGGCGCG 215
170 ATATCGTCGGTGTATCCGTACCGACGATGAGCGCGGCAAGCGTTCA----- 217
216 ACATCGTCGGCATCACAGTGCAGGCAAGAGGAGGCTGAGGCGTGCATGAAGATTCCGG 275
218 ---AATCATCAACAGCAGGTTAAAGTGCAGGCGTGCAGTGCATCCACTTCGACTATC 274
276 AGCAGCTGTTCAAGGACAAGTAGCAGTGCAGGCGTGCAGTGCATCCACTTCAGCCCA 335
275 GCATTCGGTGAAGTAGGCGAATACGGGCTGCAGTGTCTGCGTATTAAACCTTGGCAATA 334
336 CGTGCCCATGATGGTGCGG---ACGCTTCGAGAAGATCCGTGTGAACCCGCGCACT 392
335 TCGGTAATG 343
393 TCGCTGACG 401

RESULT 10
AV628026
LOCUS
DEFINITION
AV628026 Chlamydomonas reinhardtii 5% to 0.04% CO2 Chlamydomonas
reinhardtii cDNA clone LCL035f07_r 5', mRNA sequence.
ACCESSION
AV628026
VERSION
AV628026.1 GI:10790660
SOURCE
Chlamydomonas reinhardtii
ORGANISM
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE
1 (bases 1 to 531)
AUTHORS
Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL
DNA Res. 7 (5), 305-307 (2000)
MEDLINE
20539644
PUBMED
11089912
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
Location/Qualifiers
1..531
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone_lib="LCL035f07_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a carbon stress acclimated condition in which carbon
dioxide concentration in the bubbling gas was changed from
5% to 0.04%"
BASE COUNT 119 a 166 c 165 g 81 t
ORIGIN
Query Match 6.0%; Score 67.4; DB 9; Length 531;
Best Local Similarity 56.6%; Pred. No. 2.3e-09;
Matches 175; Conservative 0; Mismatches 116; Indels 18; Gaps 2;

50 TTGGGAATGTGCGGATTTGGCGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGC 109
106 TCGCAGGTGCGGTTGGCAGCAGCAGCAGCAGCAGTCCCTTCAGACCATGACCAACCG 165
110 GTACGACAGCGTGAAGCAACCGTCAATCAATCAAGCGCTGGAACCGGTTGGCGCTG 169
166 ACACCCGCAATGTGACGTGACCGTGACCGAGTGAAGAGTGCAGCGGCGCGCG 225

```

```

170 ATATCGTCGGTGTATCCGTACCGACGATGAGCGGCAAGCGTTCA----- 217
226 ACATCGTCGGCATCACAGTGCAGGCAAGAGGAGGCTGAGCGTGCATGAAGATTCCGG 285
218 ---AATCATCAACAGCAGGTTAAAGTGCAGGCGTGCAGTGCATCCACTTCGACTATC 274
286 AGCAGCTGTTCAAGGACAAGTAGCAGTGCAGGCGTGCAGTGCATCCACTTCAGCCCA 345
275 GCATTCGGTGAAGTAGGCGAATACGGGCTGCAGTGTCTGCGTATTAAACCTTGGCAATA 334
346 CGTGCCCATGATGGTGCGG---ACGCTTCGAGAAGATCCGTGTGAACCCGCGCACT 402
335 TCGGTAATG 343
403 TCGCTGACG 411

RESULT 11
AV644087
LOCUS
DEFINITION
AV644087 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
cDNA clone HCL081e11_r 5', mRNA sequence.
ACCESSION
AV644087
VERSION
AV644087.1 GI:10787415
SOURCE
Chlamydomonas reinhardtii
ORGANISM
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE
1 (bases 1 to 531)
AUTHORS
Asamizu,E., Miura,K., Kucho,K., Inoue,Y., Fukuzawa,H., Ohyama,K.,
Nakamura,Y. and Tabata,S.
TITLE
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL
DNA Res. 7 (5), 305-307 (2000)
MEDLINE
20539644
PUBMED
11089912
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
Location/Qualifiers
1..531
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone_lib="HCL081e11_r"
/notes="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"
BASE COUNT 120 a 168 c 162 g 81 t
ORIGIN
Query Match 6.0%; Score 67.4; DB 9; Length 531;
Best Local Similarity 56.6%; Pred. No. 2.3e-09;
Matches 175; Conservative 0; Mismatches 116; Indels 18; Gaps 2;

50 TTGGGAATGTGCGGATTTGGCGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGC 109
97 TCGCAGGTGCGGTTGGCAGCAGCAGCAGCAGTCCCTTCAGACCATGACCAACCG 156
110 GTACGACAGCGTGAAGCAACCGTCAATCAATCAAGCGCTGGAACCGGTTGGCGCTG 169
157 ACACCCGCAATGTGACGTGACCGTGACCGAGTGAAGAGTGCAGCGGCGCGCG 216
170 ATATCGTCGGTGTATCCGTACCGACGATGAGCGGCAAGCGTTCA----- 217
217 ACATCGTCGGCATCACAGTGCAGGCAAGAGGAGGCTGAGGCGTGCATGAAGATTCCGG 276

```



```
FEATURES
  source
    Class: shotgun.
    Location/Qualifiers
      1..1268
        /organism="Pseudomonas aeruginosa"
        /mol_type="genomic DNA"
        /strain="2-164"
        /db_xref="taxon:287"
        /clone="pacs2-164.970"
        /clone_lib="pacs2-164"
        /note="clinical isolate 2-164 Whole genomic shotgun
        library."
BASE COUNT    257 a   386 c   351 g   273 t       1 others
ORIGIN
Query Match      6.1%; Score 68.2; DB 29; Length 1268;
Best Local Similarity 58.8%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 93; Indels 3; Gaps 1;
QY 853 AACGCGCTGGAGCAAGCCCTGGAAGATATCATCTCCGATGGACGTTTCGATTATCGGC 912
Db 752 AAGCACTGGAGGGCGCTGGAGACCTCTGATCCGATGACGCTGGCGGTGATCGGT 693
QY 913 TGGTGGTGAATGCCAGGTAGGCGCTGGTTTCTACATCTCGGCTACCGGCGGCAAC 972
Db 692 TGGTGGTCAACGGTCCGTGCGAAGCAAGAGGCGCATGTCTGGCGCTCACCGGCTGCACT 633
QY 973 AAGAAAGCGGCTCTATGAAGATGCGTGGCGAAGACCGTCTGGACACACGATATG 1032
Db 632 CCGAA---CCTGGTGTATATCGACGGAAGCCGCTCGCAGAACTGACCAACGAACTTG 576
QY 1033 ATGACCACTGGAGCAGCAGCTTCCTGGGAAAGCCAGTCAGCTGGACGAGC 1085
Db 575 GTGACGACGTGGAAACGGTGTATCCCGCAGAGCGGCCGAGAGCCCGAGGC 523

RESULT 6
AV644003
LOCUS
DEFINITION
  AV644003 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
  CDNA clone HCL080a02_f 5', mRNA sequence.
ACCESSION
  AV644003
VERSION
  AV644003.1 GI:10787331
KEYWORDS
  EST.
SOURCE
  Chlamydomonas reinhardtii
  ORGANISM
    Chlamydomonas reinhardtii
    Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
    Chlamydomonadaceae; Chlamydomonas.
REFERENCE
  1 (bases 1 to 448)
  Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
  Nakamura, Y. and Tabata, S.
  Generation of expressed sequence tags from low-CO2 and high-CO2
  adapted cells of Chlamydomonas reinhardtii
  DNA Res. 7 (5), 305-307 (2000)
PUBMED
  11089912
COMMENT
  Contact: Erika Asamizu
  The First Laboratory for Plant Gene Research
  Kazusa DNA Research Institute
  Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
  Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
  Location/Qualifiers
    1..448
      /organism="Chlamydomonas reinhardtii"
      /mol_type="mRNA"
      /strain="C9"
      /db_xref="taxon:3055"
      /clone="HCL080a02.r"
      /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
      XhoI; The cDNA library was constructed from cells cultured
      in a medium with bubbling air containing 5% carbon
      dioxide"
BASE COUNT    106 a   142 c   132 g   68 t
ORIGIN

FEATURES
  source
    Class: shotgun.
    Location/Qualifiers
      1..1268
        /organism="Pseudomonas aeruginosa"
        /mol_type="genomic DNA"
        /strain="2-164"
        /db_xref="taxon:287"
        /clone="pacs2-164.970"
        /clone_lib="pacs2-164"
        /note="clinical isolate 2-164 Whole genomic shotgun
        library."
BASE COUNT    257 a   386 c   351 g   273 t       1 others
ORIGIN
Query Match      6.1%; Score 68.2; DB 29; Length 1268;
Best Local Similarity 58.8%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 93; Indels 3; Gaps 1;
QY 853 AACGCGCTGGAGCAAGCCCTGGAAGATATCATCTCCGATGGACGTTTCGATTATCGGC 912
Db 752 AAGCACTGGAGGGCGCTGGAGACCTCTGATCCGATGACGCTGGCGGTGATCGGT 693
QY 913 TGGTGGTGAATGCCAGGTAGGCGCTGGTTTCTACATCTCGGCTACCGGCGGCAAC 972
Db 692 TGGTGGTCAACGGTCCGTGCGAAGCAAGAGGCGCATGTCTGGCGCTCACCGGCTGCACT 633
QY 973 AAGAAAGCGGCTCTATGAAGATGCGTGGCGAAGACCGTCTGGACACACGATATG 1032
Db 632 CCGAA---CCTGGTGTATATCGACGGAAGCCGCTCGCAGAACTGACCAACGAACTTG 576
QY 1033 ATGACCACTGGAGCAGCAGCTTCCTGGGAAAGCCAGTCAGCTGGACGAGC 1085
Db 575 GTGACGACGTGGAAACGGTGTATCCCGCAGAGCGGCCGAGAGCCCGAGGC 523

RESULT 7
AV642898
LOCUS
DEFINITION
  AV642898 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
  CDNA clone HCL060b04_r 5', mRNA sequence.
ACCESSION
  AV642898
VERSION
  AV642898.1 GI:10786226
KEYWORDS
  EST.
SOURCE
  Chlamydomonas reinhardtii
  ORGANISM
    Chlamydomonas reinhardtii
    Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
    Chlamydomonadaceae; Chlamydomonas.
REFERENCE
  1 (bases 1 to 482)
  Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
  Nakamura, Y. and Tabata, S.
  Generation of expressed sequence tags from low-CO2 and high-CO2
  adapted cells of Chlamydomonas reinhardtii
  DNA Res. 7 (5), 305-307 (2000)
PUBMED
  11089912
COMMENT
  Contact: Erika Asamizu
  The First Laboratory for Plant Gene Research
  Kazusa DNA Research Institute
  Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
  Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
  Location/Qualifiers
    1..482
      /organism="Chlamydomonas reinhardtii"
      /mol_type="mRNA"
      /strain="C9"
      /db_xref="taxon:3055"
      /clone="HCL060b04.r"
      /note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
      XhoI; The cDNA library was constructed from cells cultured
      in a medium with bubbling air containing 5% carbon
      dioxide"
BASE COUNT    112 a   150 c   147 g   73 t
ORIGIN
Query Match      6.0%; Score 67.4; DB 9; Length 482;
Best Local Similarity 56.6%; Pred. No. 2.1e-09;
```



```
source
1. .l323
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="WSH"
/db_xref="taxon:287"
/clone="msh2 4614"
/clone_lib="msh"
/notes="Environmental isolate. Whole genomic shotgun library."
240 a 394 c 318 g 371 t

BASE COUNT
ORIGIN
Query Match 28.5%; Score 319.4; DB 29; Length 1323;
Best Local Similarity 72.0%; Pred. No. 9.7e-87;
Matches 430; Conservative 0; Mismatches 166; Indels 1; Gaps 1;

QY 361 GTGGTTGACTGTGCGCGCATAAAAATTCGGATCCGTTATTGGCGTTAAACGCCGGATCG 420
DB 104 GTGGTCGACGCGCGCGGAGCAACATCCCGATCCGTTATCGGCTCAATGCCGGTTCCG 163
QY 421 CTGGAAGAAGATCTCAAGAAAGTATGGGAACCGACCGCCGAGCGGTTGCTGNAATCT 480
DB 164 CTGGAAGAAGACCTGCAGAGAATAACGGCAACCGACCCCGGAACCCCTGCTCGAATCG 223
QY 481 GCCATCGCTCATGTCATCTCGATCCGCTGAATTCGATCAGTTCAAAGTCAGCGTG 540
DB 224 GCCATCGCGCAGTCGATCATCTCGAAGTGGATTCAGAACTTCAAGTCAACGCTC 283
QY 541 AAAGCGTCTGACGTTCTTCCTCGCTGTGTGATCTTATCGTTTGTGCAAAACAGATCGAT 600
DB 284 AAGGCTCTCGACGTTCTTCATGGCGTCCGCGCTATCGCTGTGCGCAGGAGATCGAG 343
QY 601 CAGCGTTGCATCTGGGATCACCGAAGCGGTGTGCGGCGAGCGGGCAGTAAATCC 660
DB 344 CAGCGCTTGCACCTAGGATCATCCGAGCGCGGCTCGCTCCGCGACGGTGAAGTCG 403
QY 661 GCCATTGTTTAGTCTGTGCTGTCTGAAGGATCGCGACACGCTGCGCGTATCGCTG 720
DB 404 CGGTTGGGCTGGGATGCTCTTGGCGGAGGAAATCGCGACACCATCGGATTCCTTG 463
QY 721 GCGGCGCATCCGTCGAAGAGATCAAGTCGGTTTCGATATTTTGAATTCGCTGCGTATC 780
DB 464 GCTGCGCATCCGTCGAGGAGATCAAGTCGGTTTCGACATCTCAAGTCCCTGCACCTG 523
QY 781 CGTTCGAGGAGATCAACTTCATCGCTGCCCGACCTGTTCGCTCAGAAATTTGATTT 840
DB 524 CGCTCCCGTGGCATCAACTTCATCGCTGCCCGAGCTGTTTCGCGGAGAACTTCGACGTG 583
QY 841 ATCGGTACGTTAAACGCTGGAGCAACGCTCGAAGATATCATCTCCGATGGACGTT 900
DB 584 GTGAAGACCATGAACAGCTGGAGGGCGCTGGAGACCTGCTGTCGATGGACGTTG 643
QY 901 TCGATTATCGGCTCGCTGTGTAATGCGCCAGGTGAGCGCTGTTTCTACACTCCGC 957
DB 644 GCGGTAATC-GTTGCGTGTAAACGTCGCGGGAAGCAAGAGGCCATGTTGTTGC 699

RESULT 2
BU004160/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Lactuca sativa
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
1 (bases 1 to 351)
REFERENCE
AUTHORS
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,

Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison
, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [R.W.Michelmore@vegmail.ucdavis.edu]
singleton, see http://cgdb.ucdavis.edu/ for details.
Plate: QGG37 row: N column: 12.
Location/Qualifiers
1. .l351
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L. serriola"
/db_xref="taxon:4236"
/clone="QGG37N12"
/lab_host="E.coli"
/clone_lib="QG EFGHJ lettuce serriola"
/notes="Vector: pBRCDNA5fiAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgdb.ucdavis.edu/
TAG_SEQ=Not found"
62 a 114 c 115 g 59 t 1 others

BASE COUNT
ORIGIN
Query Match 7.3%; Score 81.8; DB 13; Length 351;
Best Local Similarity 53.1%; Pred. No. 6.7e-14;
Matches 173; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 429 AGATCTCGAAGAAAAAGTATGGCAACCGACGCGGCGGTGCTGGAAATCTGCCATGCG 488
DB 332 AGCTCTCGCCCAAGTAGCGGCGCGCCACGCTGAGCGGCTCGTGGCTCGCTCTCAA 273
QY 489 TCATGTTGATCATCTGATCGCTGAATTCGATCAGTTCAAGTCAGCTGAAACGGTC 548
DB 272 GGAAGCGCGCATGTTTCGAAGATCTGGGCTTCCCGCTTCAAGATCTCGTCAAGCACA 213
QY 549 TCACGCTCTTCCTCGCTGTTGAGTCTTATCGTTTGTGCAAAACAGATCGATCAGCGCTT 608
DB 212 CGGCCCGGTACACATGTTGGAGACCTACCGATCTCGCTCTCAAGGAGACTGTCCCT 153
QY 609 GCATCTGGGGATCACCGAAGCGGTGGTGGCGGCGAGCGGGGAGTAAATCCGCCATTGG 668
DB 152 GCACCTGGAGTCACTGAGCAGGTCGCGCTGGCAGGCCCCCATCAAGTCTTGGCTGGC 93
QY 669 TTTAGTCTGCTGCTGTTCTGAAGGATCGGCACACGCTGCGGCTATCGCTGCGCGCGGA 728
DB 92 ATTGCGTGGCTGCTGCGCGAAGGATCGGCACACGATTTGGGGTCTTCTCTCNCGCGC 33
QY 729 TCCGCTCGAAGAGATCAAGTCGTT 754
DB 32 TCTGTGGAGAGGTGAAGTGGGTT 7

RESULT 3
BH235010
LOCUS
DEFINITION
ACCESSION
VERSION
BH235010.1 GI:18030478
BH235010
720 bp
DNA
linear
GSS 01-JAN-2002
PH_05.x Spiroplasma kunkelii H Spiroplasma kunkelii genomic clone
PH_05.x, genomic survey sequence.
BH235010
ACCESSION
BH235010.1 GI:18030478
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 13:42:06 ; Search time 2752 Seconds

(without alignments)  
9882.525 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgataaccaggctccaat.....ttcagcagggttgaaaaataa 1119

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estnu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_est3:\*

12: gb\_est4:\*

13: gb\_est5:\*

14: gb\_est6:\*

15: em\_estfun:\*

16: em\_estom:\*

17: em\_gss\_hum:\*

18: em\_gss\_inv:\*

19: em\_gss\_pln:\*

20: em\_gss\_vrt:\*

21: em\_gss\_fun:\*

22: em\_gss\_mam:\*

23: em\_gss\_mus:\*

24: em\_gss\_pro:\*

25: em\_gss\_tod:\*

26: em\_gss\_phg:\*

27: em\_gss\_vrl:\*

28: gb\_gss1:\*

29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	319.4	28.5	1323	29	BZ575759 msh2_4614
2	81.8	7.3	351	13	BU004160 QGG37N12
3	81.6	7.3	720	28	BH235010 PH_05.x S
4	74.8	6.7	1127	29	BZ577824 msh2_5589

C	5	68.2	6.1	1268	29	BZ569685
	6	67.8	6.0	448	9	AV644003
	7	67.4	6.0	482	9	AV642898
	8	67.4	6.0	502	9	AV626792
	9	67.4	6.0	512	9	AV644476
	10	67.4	6.0	531	9	AV628026
	11	67.4	6.0	531	9	AV644087
	12	67.4	6.0	545	9	AV626844
	13	65.8	5.9	467	9	AV629427
	14	62	5.5	377	9	AV641168
	15	58	5.2	688	14	CD424926
	16	58	5.2	760	14	CD423530
	17	56.8	5.1	792	14	CB620208
	18	56.8	5.1	821	14	CB628414
	19	56.8	5.1	852	14	CB628479
	20	56.6	5.1	658	14	CD233378
	21	56.4	5.0	380	9	AI770308
	22	54.2	4.8	303	9	AV640429
	23	53.8	4.8	597	9	AW282425
	24	53.6	4.8	728	10	BG525868
C	25	53.4	4.8	658	28	AZ570993
	26	52.6	4.7	640	9	AW738440
	27	52.6	4.7	680	12	BG890681
	28	51.8	4.6	731	10	BG124857
	29	51.6	4.6	748	14	CB662233
	30	51.6	4.6	770	14	CB662235
	31	51.6	4.6	828	14	CB669733
	32	51.6	4.6	853	14	CB619369
	33	51.4	4.6	492	12	BM323500
	34	51.4	4.6	580	14	CD423350
	35	51.4	4.6	722	14	CA483211
	36	51	4.6	579	12	BI935552
	37	51	4.6	700	12	BI933447
	38	50.4	4.5	173	10	BF962770
	39	50	4.5	662	14	CB662234
	40	49.8	4.5	576	13	BQ975044
	41	49.6	4.4	489	6	AUI86794
	42	49.4	4.4	532	10	BE924238
	43	49.4	4.4	645	12	BI934498
	44	48.4	4.3	558	9	AV917069
	45	48.4	4.3	603	12	BJ551332

ALIGNMENTS

RESULT 1	BZ575759	1323 bp	DNA	linear	GSS 17-DEC-2002
LOCUS	msh2_4614.x1	msh	Pseudomonas aeruginosa	genomic clone	msh2_4614,
DEFINITION	genomic survey sequence.				
ACCESSION	BZ575759				
VERSION	BZ575759.1	GI:27210820			
KEYWORDS	GSS.				
SOURCE	Pseudomonas aeruginosa				
ORGANISM	Pseudomonas aeruginosa				
REFERENCE	1	(bases 1 to 1323)			
AUTHORS	Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.				
TITLE	Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library				
JOURNAL	J. Bacteriol., (2002) In press				
COMMENT	Contact: Chris K. Raymond Genome Center University of Washington Box 352145, Seattle, WA 98105-2145, USA Tel: 2062216954 Fax: 2066857244 Email: craymond@u.washington.edu Class: shotgun.				

FEATURES Location/Qualifiers

**THIS PAGE BLANK (USPTO)**

• : 1-4\*

Mon Nov 24 16:22:43 2003

us-09-921-992-3.rng

Page 15

Job time : 372 secs

---





Db 552 TTATGCAATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAAATATCTCTGCATCTCGG 611  
Qy 618 GATCACCAGGCGGCGGCGAGCGGCGAGTAAATCCGCAATCGCTTGGTGTCT 677  
Db 612 AATTACAGAACTGGGTACAAATTTGCTGGAGGAATAAAAGTCTGCTGTTTAGGAGC 671  
Qy 678 GCTGCTGTCTGAAGGCGATCGGCACACGCTGCGCGTATCGCTGGCGCGGATCCGCTCGA 737  
Db 672 GATACTCAGTTTGGGCATTGGAAATACATTAACGAGTATCTTTGAGTCTGATCTGTGA 731  
Qy 738 AGAGATCAAGTCGCTTTCGATATTTTGAATCGCTGCGTATCCGTTCCGAGGAGTCAA 797  
Db 732 AGAAATAAAGTGGCGCGGGAAGTTTAAATCATTTGGTCTTTCCCTCGAATGCCCAAT 791  
Qy 798 CTTTCATCGCTGCCGACCTGTTGCGGTGAGGATTTGATGTTATCGGTACGTTAACGC 857  
Db 792 GCTTATCTCTGCTGCTTACCTGCGGTGGAATAGAGATTTGAATTAATCGTATCGCTAATGA 851  
Qy 858 GCTGGAGCAACGCTCGAAGATATCATCTCCGATGGAGCTTTGATTTATCGGCTGCGT 917  
Db 852 AGTGGAAATACATAGCAAGATTGAAGTTCGATTAAGTAGCGTCTGGCTGTGC 911  
Qy 918 GGTGAATGCCAGGTGAGCGCTGTTTCTACACTCGCGGTACCGCGGCAACAAAGAA 977  
Db 912 GGTCAACGCGCTGGAGAGCTCGGAAGCCGATATCGGAATGCTGTTCAACACGAGA 971  
Qy 978 AAGCGCCTCTATGAAGATGG 998  
Db 972 AGGCCTTCTTTTAGACATGG 992

## RESULT 14

ABQ69993  
ID ABQ69993 standard; DNA; 1107 BP.

AC ABQ69993;

DT 29-AUG-2002 (first entry)

DE Listeria monocytogenes EGDe DNA sequence #205.

KW Antibacterial; Listeria; food contamination; mutational analysis;

XX infection; ds.

OS Listeria monocytogenes EGDe.

XX WO200228891-A2.

PN 11-APR-2002.

PD 04-OCT-2001; 2001WO-FR03061.

PF 04-OCT-2000; 2000FR-0012697.

PR (INSP ) INST PASTEUR.

PA (CNRS ) CNRS CENT NAT RECH SCI.

XX Kunet F, Glaser P;

PI WPI; 2002-332479/37.

DR New genomic sequences from Listeria species, useful for detection,

XX treatment and prevention of infection, also related polypeptides,

PT antibodies and modulators -

XX Claim 16; SEQ ID 2806; 180pp; French.

XX The present invention relates to nucleic acid sequences

CC (ABQ67188-ABQ7121) from Listeria sp. The sequences are useful as probes

CC and primers for identification and/or detection of Listeria (e.g. as

CC used to screen for compounds that modulate gene expression, replication

CC and pathogenicity of Listeria (potential therapeutic agents), also for

CC treating infections by Listeria, and are useful as immunogens in

CC anti-Listeria vaccines.

CC Note: the sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic format

CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Query Match 22.5%; Score 251.4; DB 24; Length 1107;

Best Local Similarity 53.5%; Pred. No. 9.7e-63;

Matches 525; Conservative 0; Mismatches 456; Indels 0; Gaps 0;

Qy 18 AATTCACGCTAGAAATCAACACGATTTTACGTTGCGAATGTCGATTTGGCGATGGTGC 77

Db 12 AATATTCGCGAAACACTCGCCAGTCCAAGTGGGTAAATTAATTTGGTGGTAGTA 71

Qy 78 TCCCATCCCGTACAGTCCATGACCAATACGCTACGACACGCTCGAAGCAACGCTCAA 137

Db 72 GGAATTAATCTCCAAAGCATGACTACTACAAGACACATGATGTCGAAGCAACAGTAGC 131

Qy 138 TCAATCAAGCGCTGGAAACGCGTTGGCGCTGATATCGTCCGTGATCCGTAACGACGAT 197

Db 132 AGAAATTCACCGATTAGAAGAAGCTGGTTGTTCAGATTGTCGAGTTGCTTGTCTGATGA 191

Qy 198 GGACGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAACGTCGCCGCTGGTGGCTGA 257

Db 192 ACGTGCAGCGAATGCTCTTAGTGCCATCAAGAAAAGATTTCATATTCGCTTTGTGGCAGA 251

Qy 258 CATCCACTTCGACTATCGCATTCGCTGCGTAAGTAGCGGAATACGCGCTCGATTGCTCGG 317

Db 252 TATTCATTTTGTATACCGACTAGCACTTAAGCTATTGATGCAGGTGTGACAAAATTGC 311

Qy 318 TATTAAACCTGGCAATATCGGTAATGAAGAGCGTATTCGCATGGTGGTGTGATGTCGCGG 377

Db 312 GATTAACTCTGTAACATTTGGTCCGCTGATCGGGTGGAAAAGTGGTTAATGCTGCTAA 371

Qy 378 CGATAAAACATTCGATCCGTTATGGCGTTAAACGCCGATCGCTGGAAAAGATCTGCA 437

Db 372 AGCAAAAATATTCCAATCCGTTATGGGGTTAACTGCTGTTAGTAAAGAAAATAATAT 431

Qy 438 AGAAAGTATGCGAACCGACCGCGCGCTGCTGGAATCTGCGCATGCGCTCATGTTGA 497

Db 432 TCAAAAATATGGTTACCTACTCTGACGCGAATGGTAGAAGTGCACTTGCCCATTAATA 491

Qy 498 TCATTCGATCCCTGAACTTCGATCAGTTCAAAAGTCAGCGTGAAGCGTCTGACGCTTT 557

Db 492 AATTCGCAAGATTTAGATTTTATGATATCATTTGTTCTTTGAAGGCTTCTGATGTGA 551

Qy 558 CTTGCTGTTGAGTCTTATCGTTTGTGCGCAAAACAGATCGATCAGCGCTGCTATCGG 617

Db 552 TTTAGCAATTGAAGCTTATGATAAGCTAGTCGCGCATTTAATATCTCTGCTGCTCGG 611

Qy 618 GATCACCAGACCGGTGTGCGCGCAGCGGCGAGTAAATCCCGCATTTGGTTTAGTCT 677

Db 612 AATTACAGAACTCTGGTACAAATTTGCTGGAGGAATAAAAAGTGTCTGCTGTTAGGAGC 671

Qy 678 GCTGCTGTCTGAAGGCATCGGCGACACGCTCGCGGTATCGCTGGCGCGCGATCCGGTGA 737

Db 672 GATACTCAGTTTGGGCATTGGAAATACATTACGAGTATCTTTAGTGTCTGATCTCTGTA 731

Qy 738 AGAGATCAAGTCGTTTCGATATTTTGAATTCGCTGCTGATCCGTTCCGAGGAGTCAA 797

Db 732 AGAAATAAAGTGGCGCGGGAAGTTTAAATCATTTTGGTCTTTCTCTCGAATGCCCAAT 791

Qy 798 CTTTCATGCTGCCGACCTGTTGCGGTACAGAAATTCATGTTATCGGTACGGTTAAACGC 857

Db 792 GCTTATCTCTGCTGCTTACTCTGCGGTGCAATAGAGATTGATTTAATTCGTATCGCTAATA 851

Qy 858 GCTGGAGCAACGCTGGGAAGATATCATCACCTCCGATGGACGTTTCGATTTATCGGCTGGGT 917

Db 852 AGTGGAAAATTTACATAGCAAGATTGAAGTTCCGATTAAAGTAGCCGCTGCTGGCTGCTGC 911







XX DE DNA encoding novel human diagnostic protein #25870.  
 XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US08631.  
 XX PR 31-MAR-2000; 2000US-0540217.  
 XX PR 23-AUG-2000; 2000US-0649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Drmanac RT, Liu C, Tang YT;  
 XX DR WPI: 2001-639362/73.  
 XX DR F-PSDB; ABG25879.  
 XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX PT diagnostics, forensics, gene mapping, identification of mutations  
 XX PT responsible for genetic disorders or other traits and to assess  
 XX PT biodiversity -  
 XX PS Claim 1; SEQ ID NO 25870; 103pp; English.  
 XX CC The invention relates to isolated polynucleotide (I) and  
 XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 XX CC and gene mapping, and in recombinant production of (II). The  
 XX CC polynucleotides are also used in diagnostics as expressed sequence tags  
 XX CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 XX CC to restore normal activity of (II) or to treat disease states involving  
 XX CC (II). (II) is useful for generating antibodies against it, detecting or  
 XX CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 XX CC a food supplement. (II) and its binding partners are useful in medical  
 XX CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 XX CC disorders involving aberrant protein expression or biological activity.  
 XX CC The polypeptide and polynucleotide sequences have applications in  
 XX CC diagnostics, forensics, gene mapping, identification of mutations  
 XX CC responsible for genetic disorders or other traits to assess biodiversity  
 XX CC and to produce other types of data and products dependent on DNA and  
 XX CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 XX CC diagnostic coding sequences of the invention.  
 XX CC Note: The sequence data for this patent did not appear in the printed  
 XX CC specification, but was obtained in electronic format directly from WIPO  
 XX CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 1857 BP; 463 A; 472 C; 530 G; 392 T; 0 other;  
 XX Query Match 31.8%; Score 356; DB 23; Length 1857;  
 XX Best Local Similarity 97.9%; Pred. No. 4.4e-93;  
 XX Matches 466; Conservative 0; Mismatches 0; Indels 10; Gaps 10;  
 QY 644 GCGGGGCGAGTAAATCCCGCATTTGGTTTAGGTCTGCTGCTGAAGGCATCGGCGACA 703  
 DB 512 GCGGGGCGAGTAAATCCCGCATTTGGTTTAGG-CTGCTGCTGCTGAGGCATCGGCGACA 570  
 QY 704 CGCTGGCGGTATCGTGGGGCGGATCCGGTCCGAAGAGATCAAGTCGGTTTCGATATT 763  
 DB 571 CGCTGGCGGTATCGTGGGGCGG-CGATCCGGTCCGAAGAGATCAAGTCGGTTTCGATATT 629  
 QY 764 TGAATCGCTGGTATCGTGGGGGATCACTTCATCCGCTGCCGACCTGTTCCG 823  
 DB 630 TGAATCGCTG-GTATCCGTTCCGAGGGATCACTTCATCCGCTGCCGACCTGTTCCG 688  
 QY 824 GTCCAGGAATTTGATGTTATCGGTACGGTTAACCGCTGGAGCAACGCCCTGGAGGATATCA 883

DB 689 G-CAGGAATTTGATGTTATCGGTACGGTTAACCGCTGGAGCAACGCCCTGG-AGATATCA 746  
 QY 884 TCACCTCCGATGACGCTTTTCGATTATCGCTGCTGTTGAATGCGCCAGGTGAGGCGCTGG 943  
 DB 747 TCACCTCCGATGACGCTTTTCGATTATCGCTGCTGTTGAAT-GCCAGGTGAGGCGCTGG 805  
 QY 944 TTTCTACACTCGGCTCACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGCGGTGC 1003  
 DB 806 TTTCTACACTCGGCTCACCGCGCGCAACA-AAAAGCGGCTCTATGAAGATG-GTGC 863  
 QY 1004 GCAAGACCGTCTGGACAACAACGATATGATCGACGCTGGAAGCACCATTCGTGCGA 1063  
 DB 864 GCAAGACCGTCTGGACAACA-GATATGATCGACGCTGGAAGCACCATTCGTGCGA 922  
 QY 1064 AGCCAGTCAGCTGCAAGCGCGCTCGAATTCAGCTTCAGCAGTTGAAAAATAA 1119  
 DB 923 AAGCCAGTCAGC-GGACGAAGCGCTCGAATTCAGCAGTTGAAAAATAA 977  
 RESULT 10  
 ABQ81842  
 ID ABQ81842 standard; DNA; 349980 BP.  
 XX AC ABQ81842;  
 XX DT 19-NOV-2002 (first entry)  
 XX DE Bifidobacterium longum NCC2705 genomic sequence SEQ ID NO:1.  
 XX KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
 KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;  
 KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
 KW rotavirus; food composition; pharmaceutical composition; gene; ds.  
 XX OS Bifidobacterium longum.  
 XX PN EP1227152-A1.  
 XX PD 31-JUL-2002.  
 XX PF 30-JAN-2001; 2001EP-0102050.  
 XX PR 30-JAN-2001; 2001EP-0102050.  
 XX PA (NEST ) SOC PROD NESTLE SA.  
 XX DR WPI; 2002-668397/72.  
 XX PT Novel polynucleotide comprising Bifidobacterium genome sequence useful  
 PT as a probe or primer for detecting and/or identifying Bifidobacterium  
 PT longum in a biological sample -  
 XX PS Claim 1; SEQ ID 1; 80pp; English.  
 XX CC The present invention describes a polynucleotide (I) comprising a  
 CC sequence of a Bifidobacterium genome selected from the nucleotide  
 CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
 CC least 90% identity or which hybridises with the sequences given in  
 CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding  
 CC a fusion protein, comprising a sequence selected from 1097 sequences  
 CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide  
 CC encoding a heterologous polypeptide. (I) has antidiarrheic and  
 CC antibacterial activities, and can be used as an inhibitor of Salmonella.  
 CC (I) (which is a probe) is useful for the detection and/or identification  
 CC of Bifidobacterium longum in a biological sample. A carrier containing  
 CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618)  
 CC can be used for preventing and/or treating diarrhoea brought about by  
 CC pathogenic bacteria and/or rotavirus. The carrier is a food composition  
 CC selected from milk, yogurt, curd, cheese, fermented milks, milk based  
 CC fermented products, ice-creams, fermented cereal based products, milk  
 CC based powders, infant formula, pet food or a pharmaceutical composition  
 CC selected from tablets, liquid bacterial suspensions, dried oral  
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.

```
QY 979 AGCGGCTCTATGAAGATGGCGT---GCGAAAGACCGTCTGGACAAACAGATATGATC 1035
DB 315253 AGTGCATTTTATGACACGGAGTGAACACAAAGAAATAAAACCGAAGAAATTATA 315312
QY 1036 GACCAGCTGGAAGCACGCATTCGTGCGAAA 1065
DB 315313 GAAAAATGGAATTAATAATTCGAAAAAAA 315342

RESULT 8
AAF28536
ID AAF28536 standard; DNA; 33140 BP.
XX
AC AAF28536;
XX
DT 04-APR-2001 (first entry)
XX
DE Genomic fragment #23.
XX
KW Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
XX
OS Moraxella catarrhalis.
XX
PN WO200078968-A2.
XX
PD 28-DEC-2000.
XX
PF 16-JUN-2000; 2000WO-US16649.
XX
PR 18-JUN-1999; 99US-0140121.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Lagace RE, Patterson C, Berg KL;
DR WPI; 2001-041427/05.
XX
PT Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PS acids -
XX
PS Claim 1; Page 191-199; 545pp; English.
XX
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX
SQ Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;

Query Match 34.2%; Score 382.6; DB 22; Length 33140;
Best Local Similarity 59.2%; Pred. No. 2.7e-100;
Matches 652; Conservative 0; Mismatches 449; Indels 0; Gaps 0;

QY 1 ATGCATAACGAGCTCAATTAACGTAGANAATCAACAGCTATTACGTTGGGAATGTG 60
DB 21735 ATGTCAATACAAATCCGATTAACCGCCCTACCAAAAAAATCTATGTTGGCAATGTC 21794
QY 61 CGGATTGGGATGCTCCCATCGCGTACAGTCCATGACCAATACGCGTAGCAGAC 120
DB 21795 GCATCGGCGGATGACCCATCAGGTCGCAAGTATGCAATACAAACACTGTGAT 21854
QY 121 GTCGAAGCAACGGTCAATCAATCAAGCGGCTGGAACGGCGTGGCGGTGATATCGTCCGT 180
DB 21855 ATTGATGCAACAGTTGCACAAATTCAGCGATGCGTGGACCGAGGTGGCGATTGATGCGT 21914
```



Db 2555 TCGGCGACACGCTGCGCGTATCGTGGCGCGGATCGGTCGAGAGATCAAAAGTCGGTT 2614  
QY 755 TCGATATTTTGAATCGCTCGGTATCGTTTCGAGGAGTCAACTTCATCGCCTGCCCGA 814  
Db 2615 TCGATATTTTGAATCGCTCGGTATCGTTTCGAGGAGTCAACTTCATCGCCTGCCCGA 2674  
QY 815 CTTGTTGCGGTACGAAATTTGATGTTATCGGTACGGTTAAACGGCTGGAGCAACGCTGG 874  
Db 2675 CTTGTTGCGGTACGAAATTTGATGTTATCGGTACGGTTAAACGGCTGGAGCAACGCTGG 2734  
QY 875 AAGATATCATCATCGATCGAGAGTTTCGATATCGCTCGGTGGTGAATGGCCAGGTG 934  
Db 2735 AAGATATCATCATCGATCGAGAGTTTCGATATCGCTCGGTGGTGAATGGCCAGGTG 2794  
QY 935 AGCGCTGGTTTCTACACTCGCGGTCAACGGCGGCAACAGAAAGCGGCTCTATGAAG 994  
Db 2795 AGCGCTGGTTTCTACACTCGCGGTCAACGGCGGCAACAGAAAGCGGCTCTATGAAG 2854  
QY 995 ATGGCGTGGCAAGACCGTCTGGCAACAACGATATGATCGACCGAGCTGGAGACGCA 1054  
Db 2855 ATGGCGTGGCAAGACCGTCTGGCAACAACGATATGATCGACCGAGCTGGAGACGCA 2914  
QY 1055 TTGTCGGAAGCCAGTCAGCTG 1077  
Db 2915 TTGTCGGAAGCCAGTCAGCTG 2937

RESULT 7  
ID ABA92787 standard; DNA; 640681 BP.  
XX ABA92787;  
AC ABA92787;  
XX ABA92787;  
DT 27-MAR-2002 (first entry)  
XX Buchnera sp. genomic DNA SEQ ID NO:1.  
XX Buchnera; cockroach-symbiotic bacterium; cockroach extermination;  
KW circular; ds.  
XX Buchnera sp.  
XX JP2001292771-A.  
XX 23-OCT-2001.  
XX 07-APR-2000; 2000JP-0107160.  
XX 07-APR-2000; 2000JP-0107160.  
XX (RIKA ) RIKAGAKU KENKYUSHO.  
XX WPI; 2002-126043/17.  
XX A genomic DNA of cockroach-symbiotic bacterium -  
XX Claim 1; Page 16-230; 237pp; Japanese.  
XX The present invention describes a gene (I) derived from Buchnera sp.  
XX containing the DNA (a) or (b), (a) has a fully defined base pair  
XX sequence selected from a table of sequences found in the Buchnera sp.  
XX genomic DNA of ABA92787 given in the specification or is a DNA selected  
XX from complementary DNA sequences, and (b) is a DNA which hybridizes with  
XX the DNA (a) and encodes a protein. Also described are: (1) a recombinant  
XX vector (II) containing (I); (2) a transformant (III) containing (II);  
XX (3) a genomic DNA of Buchnera sp. containing the sequence given in  
XX ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or  
XX (d), (c) is a DNA containing a fully defined sequence given in ABA92788  
XX or ABA92789 and (d) is a plasmid which hybridizes with a DNA; and (5) a  
XX method for the preparation of a protein in which (III) is cultured and  
XX the expression protein of the objective protein is collected from the  
XX resultant culture. The DNA is useful for developing agricultural  
XX chemicals for exterminating cockroaches. The present sequence represents

CC the specifically claimed Buchnera sp. genomic DNA sequence, from the  
CC present invention.  
XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;  
SQ  
Query Match 36.1%; Score 403.6; DB 24; Length 640681;  
Best Local Similarity 62.2%; Pred. No. 7.5e-106;  
Matches 653; Conservative 0; Mismatches 394; Indels 3; Gaps 1;  
QY 19 ATTCAACGTAGAAAATCAACACGTTATTTACGTTTGGGAATGTCCGATTTGGCGATGTGCT 78  
Db 314293 ATCAATAGAGAATAATCTGATCGTTATTTATTTGGAAAAAGTCGTTATTTGGCAATAATGCG 314352  
QY 79 CCATCGCGGTACAGTCCATGACCAATACGCTAGCAGACAGCTGCAAGCAACGGTCAAT 138  
Db 314353 CCAATATCAGTTCAATCTATGACAAATATCTCGTACTAATAATCTCTGAAATATTAAT 314412  
QY 139 CAAATCAAGGCGCTGGAAACGCTTGGCGCTGATATCGTCCGTGTATCCGTACCGACGATG 198  
Db 314413 CAAATCTTAGAGTTACAAAAGTAGGAGTAGATATTTGTCGTATTTCTATACCAATTTA 314472  
QY 199 GACGCGCAGAAAGCGTTTCAAACCTCATCAAACAGCAGGTTAAACGTCGCCGCTGGTGGCTGAC 258  
Db 314473 AAAGCTGCAGAAATCATTTCAAGAAATATAAAACAAACAAATGTTCCATTGATTGCAGAT 314532  
QY 259 ATCCACTTCGACTATCGCATTCGGCTGAAAGTAGCGGATACGGCTGCATTTGCTCGCT 318  
Db 314533 ATACATTTTGATTACAGATTAGCTTTTACAAGCTATAAATAATGGTGCAATTTGTTTAAAGA 314592  
QY 319 ATTAACCCCTGGCAATATCGGTAATGAAGAGCGTATTTGCGATGGTGTGACTGTGCGCG 378  
Db 314593 ATTAATCTGGGAATATTGGAATAAAGAAAGATATCAGAATCATTTCTTACGCAAAA 314652  
QY 379 GATAAAACATTTCCGATCCGTTATTCGGTTAAACGGGATTCGCGGAAAGATCTGCAAA 438  
Db 314653 GATGAAAATATTCCAAATTCGTTATGTTGTTAATGCTGGATCTTTAGAAAAAGATATATTA 314712  
QY 439 GAAAAGTAGTGGCAACCGACGCGGAGCGGTGCTGGAAATCTGCCATCGCTCATGTTGAT 498  
Db 314713 AAAAAATATAAATAACCTACTCCAGATGCATTAGTAGAATACAGTATGAGGCATATTGAA 314772  
QY 499 CATCTCGATCGCTGAACTTCGATCAGTTTCAAGTCAGCGTGAAGCGCTCTCACGCTCTTC 558  
Db 314773 TACTTTGATGCTTTAAATTTTAAATCAATTTAAAGTTAGTTTAAAGCGTCTCATGTATT 314832  
QY 559 CTCGCTGTGAGTCTTATCGTTTGTGCGCAACAGATCGATCAGCCGTTGCGATCTGGGG 618  
Db 314833 TTAGCTATTGAATCATATCGCATGTTAGGAAAAAATAACACAACCTTTGCATATTGGC 314892  
QY 619 ATCCCGAAGCGGTGGTGGCGGACGCGGCGGATCGCTGGCGGCGGATCGGCTGCTG 678  
Db 314893 ATAACTGAATCCGGTGGTTTAAAGGAATGGAACAGTTAATCATCTATAGGTATTGCTTTA 314952  
QY 679 CTGCTGTCTGAAGCATCGCGGACACGCTGGCGGTATCGCTGGCGGCGGATCGGCTGCGAA 738  
Db 314953 TTATTTTAGAAGCATTTGGAGATACATACGAGTTTCAATAGCGGCATCTCAACTGAA 315012  
QY 739 GAGATCAAAGTCGTTTCGATATTTTGAATTCGCTGCGTATCCGTTCCGAGGATCAAC 798  
Db 315013 GAAGTAAAGTAGGTTATGACATTTTAAAGTTTATCTTTAAAGAGCAAGAGGTATTAT 315072  
QY 799 TTCATCGCTCCCGACCTGTCGGCTCAGGAATTTGATGTTATCGGTACGGTTAACGG 858  
Db 315073 TTTATTTGCTGCTACTTTGTTCTAGACAAGAAATTTGATTAATATACAGTAAATCAA 315132  
QY 859 CTGAGAACCGCTCGGAAGATATCATCTCCGATGGAGCTTTTCCATTTATCGGTCGCTG 918  
Db 315133 CTAGAAAAAATCTAGAGATATCTCGACTCCCATAGATGATCAATTTATTTGTTGCGTT 315192  
QY 919 GTGAATGCCCGAGGTGAGCGCTGGTTTCTACATCGCGGTCAACCGCGGCAACAGAAA 978  
Db 315193 GTTAATGAATAGGTGAATCTAAAAATAGCAACTTTAGGTCTGACGAGGAAGTCATAGAAA 315252



```
QY 10 CAGGCTCCAAATTAAGTGTAGAAAATCAACACGTTATTTACGTTGGGAATGTCGCGATTGCG 69
DB 390972 CAGCCAACTATTAAAGCTGCTGATCGACAAAATTTATGTGGAAATGTACCAATTGCT 391031
QY 70 GATGTGCTCCCATCCCGGTACAGTCCATGACCAATACCGGTACGACAGACGTCGAAGCA 129
DB 391032 GGGGATGCGCTATTGCGGTGCAATCAATGACAAATACTCGACCACTGATGTGGAAGCG 391091
QY 130 ACGETCAATCAATCAAGCGGTGGAACGGTTGGCGTGTATATCGTCGTTATCCGTA 189
DB 391092 ACAGTTGCTCAAAATTAATCATTTAGAACGTTGTTGGTCAGATATTTCTGCTGATCTGT 391151
QY 190 CCGACGATGAGCGCGCAGAGCGTTCAAACTCATCAACAGCAGGTAAACGTTGCGCGTGTG 249
DB 391152 CCAACATGATGCTCGGAAGCATTTAAACAAATTAACAAAGTGAATGTTCCGCTC 391211
QY 250 GTGGCTGACATCCACTTCGACTATCGATTCGCTGAAAGTACGCGAATACGCGCTCGAT 309
DB 391212 GTAGCAGATATTCAATTCGACTATCGTATCGCGTTAAAGTCGACAAATATGGAGTGGAT 391271
QY 310 TGTCTCGCTATTAAACCTGCAATATCGGTATGAAGAGGTTATTCGATGTTGTTGAC 369
DB 391272 TGTTTACGTATCAATCTCGCAACATTTGCTGTAAGATCGCGTCCGCTGCTGTTGAT 391331
QY 370 TGTGCGCGGATAAACAACTTCGATCCGTTATGCGTTAAACCGCGATCGCTGGAATAA 429
DB 391332 TGTGCGCGGAGCAAAAATATTCGATTCGATTTGTTGTTAAATGACAGCTCTTTAGAAAAA 391391
QY 430 GATCTGCAAGAAAGTATGCGCAACCGACGCGCGGTTGCTGGAATCTGCGATCGCT 489
DB 391392 GATTTGCAAGAAATATGCGCAACCAACCGCAGAACCTTTGTTAGAACTCCGATTCGCT 391451
QY 490 CATGTTGATCATCTGATCCCTGAACTTCGATCAGTTCAAAGTACGCTGAAAGCTGCT 549
DB 391452 CATGTAGAAATCTAGATCGCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAGCGCTCC 391511
QY 550 GAGCTCTCTCGCTGTGAGTCTTATCGTTTCTGCAAAACAGATCGATCAGCGCTTG 609
DB 391512 GATGTATCTTAGCGTTGAACTTATCGTTTACTGGCTTAAGCAATTAACAGACCTTTA 391571
QY 610 CATCTGGGATCAACGAAGCGGTGTGCGCGCAGCGGGGAGTAAATCCCGCATTTGCT 669
DB 391572 CATTTAGGCATTAACAAGCAGGTGCGCAGCGCTGGTGCAGTAAATCTGCAGTGGGT 391631
QY 670 TTAGTCTGCTGCTGCTGAGGATCGCGCAGCGCTGCGGTATCGCTGCGCGCGGAT 729
DB 391632 TTAGGAATGTTATTAGCTGAGGCAATTTGGCGATACACTACGCGTCTCTTTGGCGGAGAT 391691
QY 730 CCGGTCGAAGAGATCAAGTCTGTTTTCGATTTTGAATTCGCTGCTATCCGTTCCGGA 789
DB 391692 CCTGTAGAGGAATCAAGTCTGTTTGTATTTTGAATCTTTACGGATTCGTTCAAGA 391751
QY 790 GGGATCAACTTCATCGCTGCCGACCTGTTTCGCGTCAGGAATTTGATGTTATCGGTACG 849
DB 391752 GGAATTAACCTTATGCTTSCCAACCTGTTCTCGCAAGAAATTTGATGTAATCGGTACA 391811
QY 850 GTTAACCGCTGAGCAACCCCTGGAAGATATCATCTCCGATGAGCGTTTCGATATC 909
DB 391812 GTAAATGCGCTAGAACCAACCCCTTGAAGATATTTATACCAATGATGATCTATTATC 391871
QY 910 GCGTGGTGTGATGCGCGAGGTGAGGCGCTGTTTCTACACTCGCGCTCACCGCGGC 969
DB 391872 GGTGTGTAGTGAATGTTCTGCGGAGGCACTCGTCTCCGATCTCGCGGTAACGGCGGT 391931
QY 970 AACAAAGAAAGCGGCTCTATGAGATGGCGTGCGC---AAAGACCGTCTGGCAACAAC 1026
DB 391932 AACAAAAGCGGTTATTATCTTGACGGAGAGCGCAAAAGAGCGGTTTTTGATACGAA 391991
QY 1027 GATATGATCGACAGCTGGAAGCAGCGATTCGTCGGAAGCC 1068
DB 391992 GATATAGTGAACCAATTAGGAAGCAAAAATTTCTGTCGGAAGTTC 392033
```

```
RESULT 6
AAS73006
ID AAS73006 standard; cDNA; 5484 BP.
XX
AC AAS73006;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #8810.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-P8DB; ABG08819.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 8810; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5484 BP; 1153 A; 1384 C; 1583 G; 1364 T; 0 other;

Query Match 36.8%; Score 411.8; DB 23; Length 5484;
Best Local Similarity 97.1%; Pred. No. 4e-109;
Matches 430; Conservative 0; Mismatches 12; Indels 1; Gaps 1;

QY 635 GTGCGCGCAGCGGGCAGTAAATCCGCAATTTAGTCTGCTGCTGCTGAAGGCA 694
DB 2496 GTGGGGCGCAGCGGAGTAAATCCCA-TGGTTAAGGTCTGCTGCTGAAGGCA 2554
QY 695 TCGGCGACACGCTGCGCGTATCGCTGCGGCGCATCCGTCGAAGAGATCAAAAGTCGTT 754
```

CC the same species, and to obtain nucleic acid homologues. gpE is also  
 CC used as or primers. The recombinant vectors are used in plant  
 CC transformation or transfection. gpE also act as markers capable of  
 CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).  
 CC gpE is also used to determine the level or pattern of expression of  
 CC the protein. The present sequence is Escherichia coli gpE gene.  
 XX  
 SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
 Query Match 100.0%; Score 1119; DB 24; Length 1119;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGATAAACCAGGCTCAATTCACGTAGAAAATCAACAGTATTACGTTGGGAATGTG 60  
 DB |||||  
 QY 1 ATGATAAACCAGGCTCAATTCACGTAGAAAATCAACAGTATTACGTTGGGAATGTG 60  
 DB |||||  
 QY 61 CCGATTGGCGATGGTCTCCATCGCGCTACAGTCCATACCAATACGCGTACGACAGAC 120  
 DB |||||  
 QY 61 CCGATTGGCGATGGTCTCCATCGCGCTACAGTCCATACCAATACGCGTACGACAGAC 120  
 DB |||||  
 QY 121 GTCGAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180  
 DB |||||  
 QY 121 GTCGAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180  
 DB |||||  
 QY 181 GTATCCGTACCGACGATGAGCGGCGAGAGCGCTTCAAACTCATCAACAGCAGGTTAAC 240  
 DB |||||  
 QY 181 GTATCCGTACCGACGATGAGCGGCGAGAGCGCTTCAAACTCATCAACAGCAGGTTAAC 240  
 DB |||||  
 QY 241 GTGCCGCTGGTGGCTGACATCCACTTTCGATTCGCAATTCGCGTGAAGTAGCGGAATAC 300  
 DB |||||  
 QY 241 GTGCCGCTGGTGGCTGACATCCACTTTCGATTCGCAATTCGCGTGAAGTAGCGGAATAC 300  
 DB |||||  
 QY 301 GCGTCGATGTCGCTGATTAACCTGGCAATATCGGTAAATGAAGAGGTTATTCGATG 360  
 DB |||||  
 QY 301 GCGTCGATGTCGCTGATTAACCTGGCAATATCGGTAAATGAAGAGGTTATTCGATG 360  
 DB |||||  
 QY 361 GTGGTTGACTGTGCGCGGATAAACATTCGATCCGTTTAAACGCGGATCG 420  
 DB |||||  
 QY 361 GTGGTTGACTGTGCGCGGATAAACATTCGATCCGTTTAAACGCGGATCG 420  
 DB |||||  
 QY 421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGAACCGACCGCGGCTTGTGGAATCT 480  
 DB |||||  
 QY 421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGAACCGACCGCGGCTTGTGGAATCT 480  
 DB |||||  
 QY 481 GCGATGCTGATGTCGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGT 540  
 DB |||||  
 QY 481 GCGATGCTGATGTCGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGT 540  
 DB |||||  
 QY 541 AAAGCGTCTGACGTCCTCGCTGTTGAGTCTTATCGTTTGGGCAAAACAGATCGAT 600  
 DB |||||  
 QY 541 AAAGCGTCTGACGTCCTCGCTGTTGAGTCTTATCGTTTGGGCAAAACAGATCGAT 600  
 DB |||||  
 QY 601 CAGCGCTGATCTGGGGATCACCGAAGCGGCTGTCGCGCAGCGGGGCGAGTAAATCC 660  
 DB |||||  
 QY 601 CAGCGCTGATCTGGGGATCACCGAAGCGGCTGTCGCGCAGCGGGGCGAGTAAATCC 660  
 DB |||||  
 QY 661 GCATTTGTTAGTCTGCTGTCGTTGAGGATCGCGGACACGCTGCGGCTATCGCTG 720  
 DB |||||  
 QY 661 GCATTTGTTAGTCTGCTGTCGTTGAGGATCGCGGACACGCTGCGGCTATCGCTG 720  
 DB |||||  
 QY 721 GCGGCGATCCGCTGGAAGATCAAGTCGTTTGGATATTTGAAATCGCTGCGTATC 780  
 DB |||||  
 QY 721 GCGGCGATCCGCTGGAAGATCAAGTCGTTTGGATATTTGAAATCGCTGCGTATC 780  
 DB |||||  
 QY 781 CGTTCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAATTTGATGT 840  
 DB |||||  
 QY 781 CGTTCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAATTTGATGT 840  
 DB |||||  
 QY 841 ATCGGTACGGTTAAACGCGTGGAGCAACCGCTGGAGATATCATCTCCGATGACGTT 900  
 DB |||||  
 QY 841 ATCGGTACGGTTAAACGCGTGGAGCAACCGCTGGAGATATCATCTCCGATGACGTT 900  
 DB |||||

QY 901 TCGATTATCGGCTGGCTGTGATGATGCCAGGTGAGCGCTGTTTCTACACTCGCGTC 960  
 DB |||||  
 QY 901 TCGATTATCGGCTGGCTGTGATGATGCCAGGTGAGCGCTGTTTCTACACTCGCGTC 960  
 DB |||||  
 QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAGATGGCGTGGCGAAAGACCGTCTGGAC 1020  
 DB |||||  
 QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAGATGGCGTGGCGAAAGACCGTCTGGAC 1020  
 DB |||||  
 QY 1021 AACACGATATGATCGACCGCTGGAAGCAGCATTCGTCGAAAGCCAGTCAGCTGGAC 1080  
 DB |||||  
 QY 1021 AACACGATATGATCGACCGCTGGAAGCAGCATTCGTCGAAAGCCAGTCAGCTGGAC 1080  
 DB |||||  
 QY 1081 GAAGCGCTGCAATTCAGCTTCAGCAGGTTGAAAATAA 1119  
 DB |||||  
 QY 1081 GAAGCGCTGCAATTCAGCTTCAGCAGGTTGAAAATAA 1119  
 DB |||||  
 RESULT 5  
 AAT42063  
 ID AAT42063 standard; DNA; 1830121 BP.  
 XX  
 AC AAT42063;  
 XX  
 DT 14-SEP-1999 (first entry)  
 XX  
 DE Haemophilus influenzae complete genome sequence.  
 XX  
 KW Genome; bacterium; Haemophilus influenzae; computer readable medium;  
 KW expression modulating fragment; regulation; gene expression; vector;  
 KW organism; open reading frame; ORF; ds.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WO9633276-A1.  
 XX  
 PD 24-OCT-1996.  
 XX  
 PF 22-APR-1996; 96WO-US05320.  
 XX  
 PR 07-JUN-1995; 95US-0487429.  
 PR 21-APR-1995; 95US-0426787.  
 PR 07-JUN-1995; 95US-0476102.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (UJO) UNIV JOHNS HOPKINS.  
 XX  
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;  
 XX WPI; 1996-485782/48.  
 DR  
 PT Haemophilus influenzae Rd genome recorded on computer readable  
 PT medium - useful for identifying commercially important nucleic acid  
 PT fragments by homology searching  
 XX  
 PS Claim 1; Page 77.2-77.1091; 1291pp; English.  
 XX  
 CC This sequence represents the complete genome sequence of the bacterium  
 CC Haemophilus influenzae strain Rd. The invention relates to a computer  
 CC readable medium (CRM) having recorded upon it the complete H. influenzae  
 CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide  
 CC sequence at least 99% identical to (I). By providing the full-length  
 CC genomic sequence in a computer readable form, it is possible to identify  
 CC commercially important nucleic acid fragments and expression modulating  
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to  
 CC regulate the expression of a nucleic acid molecule. Vectors and altered  
 CC organisms comprising the predicted ORFs can be used to produce any of the  
 CC polypeptide fragments of the H. influenzae Rd genome.  
 XX  
 SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;  
 Query Match 55.6%; Score 622; DB 17; Length 1830121;  
 Best Local Similarity 74.8%; Pred. No. 3.2e-169;  
 Matches 794; Conservative 0; Mismatches 265; Indels 3; Gaps 1;

CC and (iii) to identify compounds that inhibit activity of gcpE, i.e.  
CC potential antibacterial, antimycotic, antiparasitic or antiviral agents  
CC for use in humans or animals, or antiviral, antiparasitic, fungicidal  
CC or herbicidal agents for agriculture.  
XX  
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
Query Match 100.0%; Score 1119; DB 22; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCATTAACAGAGCTCAATCAAACTAGAAATCAACACGATATTACGTTGGGAATGTG 60  
Db 1 ATGCATAACAGAGCTCAATCAAACTAGAAATCAACACGATATTACGTTGGGAATGTG 60  
QY 61 CGGATTTGGGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC 120  
Db 61 CGGATTTGGGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC 120  
QY 121 GTCGAAGCAACGCTCAATCAAACTAGAAAGCGCTGGAACGGTTGGCGCTGATATCGTCCGT 180  
Db 121 GTCGAAGCAACGCTCAATCAAACTAGAAAGCGCTGGAACGGTTGGCGCTGATATCGTCCGT 180  
QY 181 GTATCGGTACCGAGTGGACGCGGAGAGCGTTTCAAACTCATCAACAGCAGGTTAAAC 240  
Db 181 GTATCGGTACCGAGTGGACGCGGAGAGCGTTTCAAACTCATCAACAGCAGGTTAAAC 240  
QY 241 GTGCCCGTGGTGGTGAATCACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 300  
Db 241 GTGCCCGTGGTGGTGAATCACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 300  
QY 301 GGGCTCGATTGCTCGGTATTAACCTGGCAATATCGGTAAATCAAGAGCGTATTCGCATG 360  
Db 301 GGGCTCGATTGCTCGGTATTAACCTGGCAATATCGGTAAATCAAGAGCGTATTCGCATG 360  
QY 361 GTGGTTGACTGTGCGCGGATATAAAACATTCGATTCGATTCGATTCGATTCGATTCGATTCG 420  
Db 361 GTGGTTGACTGTGCGCGGATATAAAACATTCGATTCGATTCGATTCGATTCGATTCGATTCG 420  
QY 421 CTGGAAAAAGATCTGAAGAAAGATATGGCAACCGGCGGCGGATTCGATTCGATTCGATTCG 480  
Db 421 CTGGAAAAAGATCTGAAGAAAGATATGGCAACCGGCGGCGGATTCGATTCGATTCGATTCG 480  
QY 481 GGCATCGCTATGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 540  
Db 481 GGCATCGCTATGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 540  
QY 541 AAAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
Db 541 AAAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
QY 601 CAGCGTTGATCTGGGATACCGAAGCGCGGTGGTGGCGGACGCGGCGGAGTAAATCC 660  
Db 601 CAGCGTTGATCTGGGATACCGAAGCGCGGTGGTGGCGGACGCGGCGGAGTAAATCC 660  
QY 661 GCCATTGTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Db 661 GCCATTGTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
QY 721 GCGGCGGATCCGCTCGAAGAGATCAAGTCCGTTTTCGATATTTTGAATCGCTCGCTATC 780  
Db 721 GCGGCGGATCCGCTCGAAGAGATCAAGTCCGTTTTCGATATTTTGAATCGCTCGCTATC 780  
QY 781 CGTTCCGAGGATCAATCTTCATCGCTGCGCACTGTTTCGCGTCAAGAAATTTGATGT 840  
Db 781 CGTTCCGAGGATCAATCTTCATCGCTGCGCACTGTTTCGCGTCAAGAAATTTGATGT 840  
QY 841 ATCGGTACCGTTAACCGCTGGAGCAACGCTCGAAGATATCATCTCCGATGGACGTT 900  
Db 841 ATCGGTACCGTTAACCGCTGGAGCAACGCTCGAAGATATCATCTCCGATGGACGTT 900  
QY 901 TCGATTATCGCTCGCTGGTGAATGGCCACGAGTGGCGGCTGTTTTCACACTCGCGCTC 960  
Db 901 TCGATTATCGCTCGCTGGTGAATGGCCACGAGTGGCGGCTGTTTTCACACTCGCGCTC 960

Db 901 TCGATTATCGCTCGCTGGTGAATGGCCACGAGTGGCGGCTGTTTTCACACTCGCGCTC 960  
QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGCTGCGCAAAAGACCGTCTGGAC 1020  
Db 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGCTGCGCAAAAGACCGTCTGGAC 1020  
QY 1021 AACACAGATATGATCGACGCTGGAAGCAGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080  
Db 1021 AACACAGATATGATCGACGCTGGAAGCAGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080  
QY 1081 GAAGCGCTGCAATTCAGCTTCAGCAGGTTGAAAAATAA 1119  
Db 1081 GAAGCGCTGCAATTCAGCTTCAGCAGGTTGAAAAATAA 1119  
RESULT 4  
AAD31203  
ID AAD31203 standard; DNA; 1119 BP.  
XX  
AC AAD31203;  
XX DT 31-MAY-2002 (first entry)  
XX Escherichia coli gcpE gene.  
DE Escherichia coli gcpE gene.  
DE  
XX gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;  
KW transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;  
KW food; feed source; transfection; single nucleotide polymorphism; SNP;  
XX oxidative stress tolerance; UV tolerance; transformation; plant; ds.  
OS Escherichia coli.  
XX  
FH Key Location/Qualifiers  
FT 1..1119  
FT /\*tag= a  
FT /product= "Escherichia coli GCPE protein"  
XX  
PN WO200212478-A2.  
XX  
PD 14-FEB-2002.  
XX  
XX 06-AUG-2001; 2001WO-US24335.  
XX  
XX 07-AUG-2000; 2000US-223483P.  
XX  
XX (MONS ) MONSANTO TECHNOLOGY LLC.  
XX  
XX Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;  
PI Valentin HE, Venkatesh TV, Venkatramesh M;  
XX  
XX WPI: 2002-227151/28.  
DR P-PSDB; AAE19653.  
XX  
XX gcpE nucleic acid which is an essential gene of the methyl-D-erythritol  
PT phosphate pathway, encoding a fully defined GCPE protein which is  
PT useful for increasing levels of tocopherol substrates in plants -  
XX  
XX Claim 3; Page 117-119; 155pp; English.  
XX  
XX The invention relates to gcpE nucleic acid molecule, an essential gene  
CC of methyl-D-erythritol phosphate (MEP) pathway that encodes rice,  
CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful  
CC for producing a transgenic plant such as Brassica campestris, B. napus,  
CC canola, castor bean, coconut, cotton, crame, linseed, maize, mustard,  
CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,  
CC or wheat with an increased isoprenoid (tocopherol) compound level. The  
CC expression of GCPE protein in organisms increases the level of  
CC tocopherol substrate such as isopentyl diphosphate and dimethylallyl  
CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE  
CC protein can nutritionally enhance food and feed sources. Overexpression  
CC of GCPE protein in transgenic plant may provide tolerance to stresses  
CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV  
CC tolerance, etc. gcpE may be used to obtain nucleic acid molecules from

CC potential in higher organisms (mostly no orthologues were identified  
CC in *Saccharomyces cerevisiae*). An antagonist or inhibitor of the  
CC expression of an essential gene or of its function provides the key  
CC for antibacterial therapy. The invention provides methods for  
CC identifying such antagonists or inhibitors. These involve  
CC contacting a bacterial cell comprising an essential gene with a  
CC candidate antagonist or inhibitor, and testing whether contact leads  
CC to cell growth inhibition and/or cell death. The method allows the  
CC development of new broad spectrum antibiotics. A conditional mutant  
CC of an essential gene can be used to induce a lethal phenotype in  
CC bacteria for the analysis of surrogate markers.

XX Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Query Match		100.0%;	Score 1119;	DB 21;	Length 1119;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1119;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGCATAACCCAGGCTCCAATTCAACGTAGAAAATCAACACGTTATTACGTTGGGAATGTG	60		
Db	1	ATGCATAACCCAGGCTCCAATTCAACGTAGAAAATCAACACGTTATTACGTTGGGAATGTG	60		
QY	61	CCGATTGGCGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC	120		
Db	61	CCGATTGGCGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC	120		
QY	121	GTGGAAGCAACCGTCAATCAATCAAGCGCTGGAACGGTTGGCGCTCATATCGTCGT	180		
Db	121	GTGGAAGCAACCGTCAATCAATCAAGCGCTGGAACGGTTGGCGCTCATATCGTCGT	180		
QY	181	GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACGAGGTTTAAAC	240		
Db	181	GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACGAGGTTTAAAC	240		
QY	241	GTGCGCGTGGTGTGACATCCACTTCGACTATCGCATTCGCGTGAAGTACGGGAATAC	300		
Db	241	GTGCGCGTGGTGTGACATCCACTTCGACTATCGCATTCGCGTGAAGTACGGGAATAC	300		
QY	301	GGCGTCGATTGTCTGGGTATTAACTCCCTGCAATATCGGTAATGAAGAGCGTATTCGCATG	360		
Db	301	GGCGTCGATTGTCTGGGTATTAACTCCCTGCAATATCGGTAATGAAGAGCGTATTCGCATG	360		
QY	361	GTGGTTGACTGTGCGCGGATAAAACATTCGATCCGTTATGGCGTTAAACCGCGGATCG	420		
Db	361	GTGGTTGACTGTGCGCGGATAAAACATTCGATCCGTTATGGCGTTAAACCGCGGATCG	420		
QY	421	CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACCGCGCGTGTGGAATCT	480		
Db	421	CTGGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACCGCGCGTGTGGAATCT	480		
QY	481	GCATGCGTCATGTTGATCATCTCGATCGCTGNACTTCGATCAGTTCAAGTCAGCGTG	540		
Db	481	GCATGCGTCATGTTGATCATCTCGATCGCTGNACTTCGATCAGTTCAAGTCAGCGTG	540		
QY	541	AAAGCGTCTGAGCTCTTCTCGCTGTGTAGTCTTATCGTTTGTGCAAAAACAGATCCGAT	600		
Db	541	AAAGCGTCTGAGCTCTTCTCGCTGTGTAGTCTTATCGTTTGTGCAAAAACAGATCCGAT	600		
QY	601	CAGCGCTGTCATCTGGGGATCAACGAACCGGTGGTGGCGGAGCGGGGAGTAAATCC	660		
Db	601	CAGCGCTGTCATCTGGGGATCAACGAACCGGTGGTGGCGGAGCGGGGAGTAAATCC	660		
QY	661	GCCATGTTTGTAGTCTGCTGTCTGAAGCATCGGACACGTCGCGGTATTCGCTG	720		
Db	661	GCCATGTTTGTAGTCTGCTGTCTGAAGCATCGGACACGTCGCGGTATTCGCTG	720		
QY	721	GCGGCGATCCCGTCAAGAGATCAAAAGTCGGTTTCGATATTTGAAATCGCTGCTATC	780		
Db	721	GCGGCGATCCCGTCAAGAGATCAAAAGTCGGTTTCGATATTTGAAATCGCTGCTATC	780		
QY	781	CGTTCGCGAGGATCAACTTCATCGCTCCCGACCTGTCGCGTCAAGAAATTTGATGTT	840		
Db	781	CGTTCGCGAGGATCAACTTCATCGCTCCCGACCTGTCGCGTCAAGAAATTTGATGTT	840		

QY	841	ATCGGTACGGTTAACGGCTGGAGCAACGCCCTGGAAGATATCATCATCCGATGACGTT	900		
Db	841	ATCGGTACGGTTAACGGCTGGAGCAACGCCCTGGAAGATATCATCATCCGATGACGTT	900		
QY	901	TCGATTATCGGCTGGTGGTGAATGCCAGGTGAGGCGCTGTTTCTACACTCGGCGTC	960		
Db	901	TCGATTATCGGCTGGTGGTGAATGCCAGGTGAGGCGCTGTTTCTACACTCGGCGTC	960		
QY	961	ACCGGCGGCACAAGAAAAGCGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC	1020		
Db	961	ACCGGCGGCACAAGAAAAGCGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC	1020		
QY	1021	AACAAACGATATGATCGACCGCTGGAAGCACCGCATTTCTGTCGAAAGCCAGTCAGCTGGAC	1080		
Db	1021	AACAAACGATATGATCGACCGCTGGAAGCACCGCATTTCTGTCGAAAGCCAGTCAGCTGGAC	1080		
QY	1081	GAAGCGCTCGAATTGACGTTTCAGCGTTGCAAGGTTGAAAAATAA	1119		
Db	1081	GAAGCGCTCGAATTGACGTTTCAGCGTTGCAAGGTTGAAAAATAA	1119		
RESULT 3					
AAC82653					
ID	AAC82653 standard; DNA; 1119 BP.				
XX	AAC82653;				
DT	15-MAR-2001 (first entry)				
DE	E. coli gcpe DNA.				
KW	Isoprenoid; gcpe; yfgB; antimicrobial; transgenic plant; agriculture;				
XX	antimycotic; antiparasitic; antiviral; fungicidal; herbicidal; ds.				
OS	Escherichia coli.				
PN	WO200072022-A1.				
PD	30-NOV-2000.				
PF	20-MAY-2000; 2000WO-EP04592.				
PR	21-MAY-1999; 99DE-1023567.				
PR	21-MAY-1999; 99DE-1023568.				
PA	(JOMA/) JOMAA H.				
PI	Jomaa H;				
DR	WPI; 2001-025196/03.				
DR	P-PSDB; AAB45692.				
PT	Incorporating gcpe and yfgB genes into viruses and cells, for				
PT	increasing isoprenoid content and identifying e.g. antimicrobial				
XX	agents, comprises using DNA sequences from bacteria or parasites				
PS	Claim 3; Page 13-15; 36pp; German.				
CC	This invention describes a novel method for incorporating gcpe and yfgB				
CC	genes into viruses and cells for increasing isoprenoid content and				
CC	identifying e.g. antimicrobial agents, comprises using DNA sequences (I)				
CC	from the gcpe or yfgB genes of bacteria or parasites or DNA sequences				
CC	(II) which hybridize to the specified genes or encode a plastid protein				
CC	with the same biological activity as those encoded by the genes. The				
CC	invention also describes (i) plant cells containing (i) or (ii); (2)				
CC	transformed plant cells, and transgenic plants regenerated from them,				
CC	that contain (I) or (II); (3) determining the enzymatic activity of a				
CC	gcpe protein; or (4) screening compounds (A) that have antimycotic,				
CC	antiparasitic or antiviral activity in humans or animals or antiviral,				
CC	antiparasitic, fungicidal or herbicidal activity in plants. (i) and (ii)				
CC	are used: (i) to increase the isoprenoid levels in viruses and cells;				
CC	(ii) for determining the enzymatic activity of gcpe and yfgB proteins;				

DR WPI; 2000-687048/67.  
XX Identifying antibacterial compounds, comprises identifying an  
PT antagonist or inhibitor of the expression of a gene encoding a  
PT polypeptide essential for bacterial growth or survival -  
XX  
PS Claim 1; Fig 1; 75pp; English.  
XX  
CC The present invention relates to antagonists and inhibitors of 24  
CC bacterial genes and proteins. The proteins are thought to be essential  
CC for growth in several species of bacteria (including S. pneumoniae, B.  
CC burgdorferi, H. influenza and H. pylori). The proteins and coding  
CC sequences shown in the specification can be used to identify antagonists  
CC and inhibitors which can be used in disease treatment and pesticides. In  
CC particular, they can be used against M. tuberculosis. The present  
CC sequence is one of the genes of the invention.  
XX  
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;  
Query Match 100.0%; Score 1119; DB 21; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCATAACACAGGCTCCAAATTCAGGTAGAAAATCAACAGCTATTACGTTGGGAATGTG 60  
DB 1 ATGCATAACACAGGCTCCAAATTCAGGTAGAAAATCAACAGCTATTACGTTGGGAATGTG 60  
QY 61 CCGATTGGCGATGGTGTCCCATCGCGGTACAGTCCATGACCAATACGCGTAGCAGAC 120  
DB 61 CCGATTGGCGATGGTGTCCCATCGCGGTACAGTCCATGACCAATACGCGTAGCAGAC 120  
QY 121 GTCGAAGCAACAGGTCATCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATATCGTCCGT 180  
DB 121 GTCGAAGCAACAGGTCATCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATATCGTCCGT 180  
QY 181 GTATCCGTACCGAGTGGACGGCGAGAAAGCGTTCAAACTCATCAACAGAGGTTAAC 240  
DB 181 GTATCCGTACCGAGTGGACGGCGAGAAAGCGTTCAAACTCATCAACAGAGGTTAAC 240  
QY 241 GTGCGCGTGGTGTGATCACTTCAGTATCGGATTCGGCTGAAAGTAGCGGATAC 300  
DB 241 GTGCGCGTGGTGTGATCACTTCAGTATCGGATTCGGCTGAAAGTAGCGGATAC 300  
QY 301 GGGTTCGATTGTCTGCGTATTACCCCTGGCAATATCGTAAATGAAGAGCGTATTTCGCATG 360  
DB 301 GGGTTCGATTGTCTGCGTATTACCCCTGGCAATATCGTAAATGAAGAGCGTATTTCGCATG 360  
QY 361 GTGTTGACTGTGCGCGCGATAAAAAATTCGATCCGTTATTGGCGTTAAACGCGGATCG 420  
DB 361 GTGTTGACTGTGCGCGCGATAAAAAATTCGATCCGTTATTGGCGTTAAACGCGGATCG 420  
QY 421 CTGGAATAAGATCTGGAAGAAAGTATGGCAACCGAGCGCGGCGTTGCTGGATCT 480  
DB 421 CTGGAATAAGATCTGGAAGAAAGTATGGCAACCGAGCGCGGCGTTGCTGGATCT 480  
QY 481 GCATGCGTCACTGTGATCATCTGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540  
DB 481 GCATGCGTCACTGTGATCATCTGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540  
QY 541 AAAGCGTCTGACGTTCTTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
DB 541 AAAGCGTCTGACGTTCTTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
QY 601 CAGCGTTGATCTGGGGATCAACGAACCGCGTGGTGGCGGACGGGGAGTAAATCC 660  
DB 601 CAGCGTTGATCTGGGGATCAACGAACCGCGTGGTGGCGGACGGGGAGTAAATCC 660  
QY 661 GCATGCTTTAGTCTGCTGCTCTGAAGGATCGGCACACGCTGCGCGTATCGCTG 720  
DB 661 GCATGCTTTAGTCTGCTGCTCTGAAGGATCGGCACACGCTGCGCGTATCGCTG 720  
QY 721 GCGGCGGATCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTGCGTATC 780  
DB 721 GCGGCGGATCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTGCGTATC 780

Db 721 GCGGCGGATCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTGCGTATC 780  
QY 781 CGTTCCGAGGATCAAACTTCATCGCTGCCGAGCTGTTCCGCTCAGGAAATTTGATGTT 840  
Db 781 CGTTCCGAGGATCAAACTTCATCGCTGCCGAGCTGTTCCGCTCAGGAAATTTGATGTT 840  
QY 841 ATCGGTACGTTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCATCCGATGGACGTT 900  
Db 841 ATCGGTACGTTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCATCCGATGGACGTT 900  
QY 901 TCGATTATCGGTGCGTGTGTAATGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTC 960  
Db 901 TCGATTATCGGTGCGTGTGTAATGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTC 960  
QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGGTGCGCAAGACCGTCTGGAC 1020  
Db 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGGTGCGCAAGACCGTCTGGAC 1020  
QY 1021 AACACGATATGATCGACCGCTGGAAGCACGCTTCGTGCGAAAGCCAGTCAGCTGGAC 1080  
Db 1021 AACACGATATGATCGACCGCTGGAAGCACGCTTCGTGCGAAAGCCAGTCAGCTGGAC 1080  
QY 1081 GAAGCGCTCGAATTGACGTTCCAGAGGTTGAAAAATAA 1119  
Db 1081 GAAGCGCTCGAATTGACGTTCCAGAGGTTGAAAAATAA 1119

RESULT 2  
AAA88705  
ID AAA88705 standard; DNA; 1119 BP.  
XX  
AC AAA88705;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE E. coli FUN essential gene gcpe.  
XX  
KW FUN gene; gcpe gene; essential gene; antibacterial; antibiotic;  
KW screening; infection; therapy; antagonist; surrogate marker; ds.  
XX  
OS Escherichia coli.  
XX  
PN BP1043403-A1.  
XX  
PD 11-OCT-2000.  
XX  
PF 09-APR-1999; 99EP-0107031.  
XX  
PR 09-APR-1999; 99EP-0107031.  
XX  
PA (GPCG-) GPC GENOME PHARM CORP AG.  
XX  
PI Not given;  
XX  
XX WPI; 2000-640125/62.  
DR  
XX Identifying antagonists of the expression of gene encoding bacterial  
PT growth polypeptide useful for treating bacterial infections or  
PT diseases, by evaluating transcription of the gene in the presence of  
PT test molecule -  
XX  
PS Claim 1; Page 25; 55pp; English.  
XX  
CC The present sequence is that of the Escherichia coli FUN gene  
CC gcpe, which encodes a protein that is essential for bacterial  
CC growth or survival. gcpe is 1 of 22 E. coli genes (see  
CC AAA88692-713) identified as being essential (there is no deletion  
CC genotype). These 22 genes fulfil criteria for being attractive  
CC antibacterial targets: hypothetical open reading frames coding for  
CC essential functions (mutation is lethal for growth in rich media);  
CC broad conservation (orthologues are present in a wide range of  
CC bacteria, including Haemophilus influenzae, Streptococcus pneumoniae,  
CC Helicobacter pylori and Borrelia burgdorferi); and low toxicity

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 11:58:06 ; Search time 358 Seconds  
(without alignments)  
8437.629 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataacagcagctccaat.....ttcagcaggtgaaataataa 1119

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_19Jun03.\*

1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*  
25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	21 AAA95478	E. coli essential
2	1119	100.0	1119	21 AAA88705	E. coli FUN essent
3	1119	100.0	1119	22 AAC82653	E. coli gcpE DNA
4	1119	100.0	1119	24 AAD31203	Escherichia coli g
5	622	55.6	1830121	17 AAT42063	Haemophilus influe
6	411.8	36.8	5484	23 AAT57306	DNA encoding novel
7	403.6	36.1	640681	24 ABA92787	Buchnera sp. genom
8	382.6	34.2	33140	22 AAF28536	Genomic fragment #

9	356	31.8	1857	23	AA990066	DNA encoding novel
10	306.2	27.4	349980	24	ABQ81842	Bifidobacterium lo
11	293.4	26.2	1083	24	ABK74401	Bacillus lichenifo
12	254.6	22.8	6157	24	ABQ70939	Listeria monocytog
13	251.4	22.5	1107	24	ABQ67966	Listeria monocytog
14	251.4	22.5	1107	24	ABQ69993	Listeria monocytog
15	251.4	22.5	2944528	24	ABA03041	Listeria monocytog
16	244	21.8	35829	23	AA595973	Propionibacterium
17	243.8	21.8	1161	25	ABZ71126	Mycobacterium tube
18	243.8	21.8	1164	22	AAH52056	Mycobacterium tube
19	243.8	21.8	4403765	22	AAI99683	Mycobacterium tube
20	243.8	21.8	4411529	22	AAI99682	Mycobacterium tube
21	243.6	21.8	994	24	ABQ68324	Listeria monocytog
22	239.4	21.4	1134	22	AAH67170	C. glutamicum codin
23	239.4	21.4	1179	25	ACA00649	C. glutamicum deri
24	239.4	21.4	349980	22	AAH68530	C. glutamicum codin
25	239.4	21.4	349980	22	AAH68531	C. glutamicum codin
26	211.8	18.9	1180	19	AAH14017	H. pylori GHPO 76
27	136.4	12.2	507	21	AACT5580	Human ORF1443 cDNA
28	136.4	12.2	507	24	ABN76496	Helicobacter pylor
29	102	9.1	435	24	ABX66784	DNA encoding novel
30	88	7.9	1152	23	AA888480	Polynucleotide seq
31	87.2	7.8	2535	20	AAZ20674	Sequence encoding
32	72.4	6.5	4467	10	AAAN92428	Zea mays partial g
33	70.6	6.3	584	24	AAAD31220	Zea mays partial g
34	68.4	6.1	596	24	AAAD31222	Zea mays partial g
35	68.4	6.1	670	24	AAAD31221	Zea mays partial g
36	58.8	5.3	1263	25	ABZ33756	N. gonorrhoeae nuc
37	58.8	5.3	60873	21	AAA81469	Neisseria meningit
38	58.8	5.3	349980	21	AAF21610	N. meningitidis B
39	58.8	5.3	1437668	21	AAA81490	Complete genome se
40	56.8	5.1	348	24	ABK78811	Bacillus clausii g
41	53.4	4.8	1038602	20	AAZ01425	Zea mays partial g
42	52.6	4.7	293	24	AAAD31224	Zea mays partial g
43	52.6	4.7	456	24	AAAD31225	Zea mays partial g
44	52.6	4.7	504	16	AAQ93779	Cross-reactive all
45	50.8	4.5	1398	20	AAZ20609	Polynucleotide seq

#### ALIGNMENTS

##### RESULT 1

AAA95478  
ID AAA95478 standard; DNA; 1119 BP.

XX AAA95478;

XX 27-FEB-2001 (first entry)

XX E. coli essential gene gcpE.

XX Bacterial growth; inhibitor; ygbB; yfhC; yacE; ychB; yefJ; ynfI;  
KW yggJ; yjeE; ylaO; yrdC; yhcB; ygbE; ygbF; yjeB; yjeC; yjeD; yjeE; yjeF;  
KW H1808; yeaA; yagF; b1983; yidD; yceG; yjbC; antibacterial compound;  
KW H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;  
KW M. tuberculosis; antibiotic; ds.  
XX Escherichia coli.  
XX WO200061793-A2.  
XX 19-OCT-2000.  
XX 07-APR-2000; 2000WO-EP03135.  
XX 09-APR-1999; 99EP-0107031.  
XX 04-FEB-2000; 2000EP-0102111.  
XX (GPCB-) GPC BIOTECH AG.  
XX Loferer H, Jacobi A;

```

; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavaram;
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 22
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Zea mays
US-09-921-992-22

```

```

Query Match      6.1%; Score 68.4; DB 9; Length 670;
Best Local Similarity 56.1%; Pred. No. 2.4e-12;
Matches 129; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 595 ATCGATCAGCGTTGTCATCTGGGGATCACCGAGCCGGTGGTGGCGCGGCGGCGTA 654
Db 223 ATGATATATCCCTCCATCTAGGGGTACCGAGCCGGGATGGGAATATGCCGCATT 282

Qy 655 AAATCCGCCATTGGTTTAGTCTGCTCTCTGAAGGCATCGCGCACACGCTGCGCGTA 714
Db 283 AAATCCACCCTGGCATTGCCACCTTTAGCTGATGGCATTGGCGATATATCCGCGTA 342

Qy 715 TCCTGGCGGCCGATCCGGTCGAAGAGATCAAGTCGGTTTCGATATTTTGAATCGCTG 774
Db 343 TCCTTCACCGAGCCCGGAAAGAAATTCGCTTTGCTACAGCATTTCTCCAGGGCGTG 402

Qy 775 CGTATCGTTTCGCGAGGATCACTTCATCGCTGCCGACCTGTTCCGG 824
Db 403 GGTTTCGGAAACCATGGTGAATATGTGGCCTGTCTTCTCTGGCCG 452

```

Search completed: November 23, 2003, 17:14:25  
Job time : 433 secs

```
Best Local Similarity 56.9%; Pred. No. 9.8e-39;
Matches 265; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 43 ATTACGTTGGCGAATGCGGATTTGGCGATGCTCCCATCCGCTACAGTCCATGACC 102
Db 233 ATCAAAATCGGGATGTTAAGATTGGGAAAAGAAATCCCGTTGTGATTCAATCGATGATT 292
QY 103 AATACCGGTACGACAGACGTCGAAGCAACCGTCAATCAATCAAGCGCTGGAACCGGTT 162
Db 293 AACACGGAATCGGATGTAGAACGAGTGTGAGGCAATTTGGATTGGGAAGAGCC 352
QY 163 GCGCTGATATCGTCCGTGTATCCGTACCGACGATGACGCGGCAAGAGGTTCAAACTC 222
Db 353 GGCTGTGAATTTGCTGATGACGATCAATACAAAAGAGCGGCTATGGCAATTCGCCGG 412
QY 223 ATCAACACGACGTTAAACGTCGCGCTGGTGTGATCCATCCATTCGACTATCGATTGG 282
Db 413 ATTAAAAGAGAGTTTCAATATCTTTTGGTAGCGGATATTCATTTGATTATCGACTTGC 472
QY 283 CTGAAGTAGCGGAATACGCGCTCGATTGCTCGTATTAAACCTGGCAATATCGGTAAT 342
Db 473 TTATTGGCGATAGAACGGAATTGCAAAATACGATTAATCCCGAAATATCGGTTCT 532
QY 343 GAAGAGCGTATTCGATGCTGTTGATGTCGCGCGCATAAAAACATTCGATCCGTAAT 402
Db 533 GAAGAAAATATTCGTTGTTGCGGAAGCGGCAAAAAGGAAGAATTCGATTCGTAAT 592
QY 403 GCGGTTAAACCGCGATCGCTGGAAGAGATCTCAAGAAAGATGTCGCAACGCGCG 462
Db 593 GGAGTCAATGACGCTTCTTTGGAAAACATATCTTGGAAAATATGGAGCGGNAACGGCA 652
QY 463 CAGCGGTTGCTGGAATCTGCATCGCTGCTGATGTTGATCATCTCGATC 508
Db 653 GATGCTNTGTCAAAAGCGCTAATGATCATGTAATAACTGTGGAAC 698

RESULT 13
US-09-921-992-21/c
; Sequence 21, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohrer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 21
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1..596)
; OTHER INFORMATION: unsure at all n locations
US-09-921-992-23

Query Match 6.3%; Score 70.6; DB 9; Length 584;
Best Local Similarity 52.5%; Pred. No. 4e-13;
Matches 179; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 595 ATCGATCAGCGGTTGCATCTGGGGATCACCGAAGCGGTGGTCGCGCAGCGGGCAGTA 654
Db 459 ATGGATTATCCCTCCATCTAGGGGTACCGAAGCGGGATGGGAATATGCCCCATT 400
QY 655 AAATCCGCCATTTGGTTAGGTCCTGCTGTCTGAAGGCGATCGGCGACACGCTGCGCGTA 714
Db 399 AAATCCACCGCTGGCATTTGCCACCTTTTAGCTGATGGCATTTGGCGATATATCCGGTA 340

RESULT 14
US-09-921-992-23
; Sequence 23, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohrer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 23
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: unsure
; LOCATION: (1..596)
; OTHER INFORMATION: unsure at all n locations
US-09-921-992-23

Query Match 6.1%; Score 68.4; DB 9; Length 596;
Best Local Similarity 56.1%; Pred. No. 2.3e-12;
Matches 129; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 595 ATCGATCAGCGGTTGCATCTGGGGATCACCGAAGCGGTGGTCGCGCAGCGGGCAGTA 654
Db 223 ATGGATTATCCCTCCATCTAGGGGTACCGAAGCGGGATGGGAATATGCCCCATT 282
QY 655 AAATCCGCCATTTGGTTAGGTCCTGCTGTCTGAAGGCGATCGGCGACACGCTGCGCGTA 714
Db 283 AAATCCACCGCTGGCATTTGCCACCTTTTAGCTGATGGCATTTGGCGATATATCCGGTA 342
QY 715 TCGCTGGGCGCGATCCGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTG 774
Db 343 TCCCTCACGAAGCGGCCGAAAAGAAATTCCTGTTGCTACAGCATTTCCAGGGCGTG 402
QY 775 CGTATCCGTTCCGAGGATCAACTTCATTCGCTGCCCGACCTGTTTCGCG 824
Db 403 GCTTTGCGGAAAACCATGCTGGAATATGTCCTTCTCTGTGGCG 452

RESULT 15
US-09-921-992-22
; Sequence 22, Application US/09921992
; Patent No. US20020069426A1
```









Qy	760	ATTTTGAATCGCTGCGTATCCGTTCCGGAGGGATCAACTTCATCGCGCTGCCGACCTGT	819
Db	3140825	ATCTGAGTGCCTGAACTTCGCGCCACGGCGGCTCGAGATCGTCTCCTGCCGTCTCGC	3140766
Qy	820	TCGCGTCAGGAATTTGATGTATTTCGTTACGTTAAACGCGCTGGAGCAACGCCCTGGGAAGAT	879
Db	3140765	GGACGCCCCAGGTTCGACGCTCTACAAGCTGCCGAGGAGTCAACGGCGGCTCTGGAGGC	3140706
Qy	880	ATCATCACTCCGATGGAAGGTTTCGATTATCGGTCGCTGGTGTGTAATGCCCCAGGTGAGCGC	939
Db	3140705	ATGAGAGTGCCCTTCGCGCTCGCGGTCATGGGGTTCGCTCGCTCAACGGCGCCGGGGAGGCC	3140646
Qy	940	CTGGTTTCTACACTTCGCGCTCACCGCGGGCAACAAGAAAGACGGCCCTCTATGAAGATGCG	999
Db	3140645	CGGAGGCCACCTTCGCGCTCGCCCTCCGCCAACGGCAAGGCACAGATCTTCTGTGAAGGCC	3140586
Qy	1000	G 1000	
Db	3140585	G 3140585	

## RESULT 7

```

US-10-156-761-1636
; Sequence 1636, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1636
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1155)
US-10-156-761-1636

```

Query Match 23.3%; Score 260.6; DB 14; Length 1155;

Query Match	23.3%	Score 200.0;	DB 14;	Length 1133;
Best Local Similarity	54.7%;	Pred. NO. 5.9e-78;		
Matches 540;	Conservative	0;	Mismatches 444;	Indels 3;
Gaps	1;			

Qy	23	AAGCTAGAAAAATCAACAGTATTACGTTTGGGAATGTCCGATTGGCGATGTGGTCTCCCA	82
Db	53	AAGCGGTGTCTCCCGGCAGATCCAGGTTCGACCGCTGGCGGTTCGGCGGCAACGCGCCCGG	112
Qy	83	TCGCCGTACAGTCCATGACCAATACGCGTTCAGACAGACGTCGAAGCAACGGTCAATCMAA	142
Db	113	TGTCCGTGCAGTCGATGACGACGCGCTAGCTCGGACATCGCGGCCACGCTCCAGCAGA	172
Qy	143	TCAAGCGCTGGAAACGCGTTGGCGCTGATATCGTCCGTGTATCCGTACCGACGATGGACG	202
Db	173	TCCGGAATCTCACCGGTCGCGCTCCAGATCGTCCGCGTCGCTGCCCTCCCAACGACGAGCG	232
Qy	203	CGCGAGAACGGTTTCAAACTCATCAAAACAGCAGGTTTACGTCGCGCTCGTGCGGTGCAATCC	262
Db	233	ACGCGACGCCCTCGCGGTATCGCCCGCAATCGCAGATCCCGCTCGTCCGCGACATCC	292
Qy	263	ACTTCGACTATCGCATTCGCTTGGCTGAAAGTAGCGGAATACGCGCTCGATTGTCTCGGTATTA	322

293	ACTTCAGCCGAAGTACGTGTTCCGCGCGATCGAGGCGGGCTGCGCGCGGTCCGGGTGA	355
323	ACCTGGCAATATCGGTAATCAAGAGCGTATTCCGATGGTGGTTGACTGTGCGGCG---	379
353	ACCGGGCAACATCAACGAGTTGACGACACAAGGTCAAGGAGATCGCGCGCGCCCAAGG	412
380	ATAAACAATTTCCGATCCGTTATTTGGCGTTTAACGCCGGATCGCTGGAAAAAGATCTGCNAG	439
413	AGCACGCGACCGCGATCCGATCGGCGTCAACGCGGCTCGCTCGACCGCGCGGTGCTGG	472
440	AAAGATATGCGACCGACGCGCCGAGCGTTGCTTGGAAATCTGCCATCGTCAATGTTGATC	499
473	AGAAGTACGGCAAGGCCACCCCGAGGCGCTGCTCGAGTCGGCGCTGTGGGAGGCGTCGC	532
500	ATCTCCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTCAGAGCGTCTGACGCTCTTCC	559
533	TCITTCGAGGAGCAGACTTCTCGGGACATCAAGATCTCGGTCAAGCAACGACCCGGTGG	592
560	TCGCTGTTAGTCTTATTCGTTTGTGCGAAAAAGATCGATCGATCGACCGTTGATCTGGGGA	619
593	TCATGGTCAACCGGTACCGCGAGCTGGCGGCGCAGTCGCATACCCGCTGCACCTCGGCG	652
620	TCACCGAAGCGGTGTCGCGCAGCGGGCAGTAAATCCGCCATTGGTTAGGTCGTCG	679
653	TGACCGAGGCGCGCCCGCTTTCAGGGCACGATCAAAGTCGGCGCTCGCCTTCGGCGCGC	712
680	TGCTGCTGAAGGATCGGCGACACGCTCGCGCGTATCGCTGGCGGCGGATCCGGTCTGAAG	739
713	TGCTCAGCGAGGACATCGGCGACACGATTCGCGTCTCGCTGAGCGCGCCCCCGTTCGAGG	772
740	AGATCAAAGTCGGTTTCGATATTTTGAATTCGCTGCGTATCCGTTCCGAGGGATCAACT	799
773	AGATCAAAGTCGGCATCCAGATCCTCGAGTCGCTGGGCCCTCAGGACGCGGCGCTGGAGA	832
800	TCATCGCCTCGCGACCTGTTCCGGTCAGGAATTTGATGTTATTCGGTACGGTTAACGCCG	859
833	TCGCTCTCTCCCGCTCTCGCGGCGCGCCAGGTCGACGCTACAAAGCTGSCCGAGGAGG	892
860	TGGAGCAACCCCTGGGAAGATATCATCACTCCGATGGAGCGTTTCGATTATCGGCTCGCTGG	919
893	TCACGGCGGGTCTGGAGGGCATGGAGTGCCCTTCGCGCTCGCGGTCATGGGCTCGCTCG	952
920	TGAATGGCCCCAGGTGAGGCGCTGTTTTCATCTCGGCGTCACCGGCGGCAACAGAAAA	979
953	TCAACGGCGCGGGAGGCCCGGAGGCCACCTTCGCGCTCGCTCCGCGCAACGCGCAAGG	1012
980	GCGGCTCTATGAAGATGGCGTCGCA	1006
1013	GACAGATCTTCGTGAAGGGCGAGGTCA	1039

## RESULT 8

```

US-09-712-363-110
; Sequence 110, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712.363
; CURRENT FILING DATE: 2000-11-13
; PRIORITY APPLICATION NUMBER: PCT/US00/02246
; PRIORITY FILING DATE: 2000-01-28
; PRIORITY APPLICATION NUMBER: 60/179,531
; PRIORITY FILING DATE: 2000-02-01
; PRIORITY APPLICATION NUMBER: 60/117,844
; PRIORITY FILING DATE: 1999-01-29
; PRIORITY APPLICATION NUMBER: 60/118,206
; PRIORITY FILING DATE: 1999-02-01
; PRIORITY APPLICATION NUMBER: 60/126,593

```



QY 919 GTGAATGCCAGGTAGGCGCTGGTTTCTACTCGGCTCAACGGCGCAACAAGAA 978  
Db 315193 GTTAATGGAATAGTGAATCTAAATAGCAACTTTAGGCTAGCAGGAAGTCATAAGAA 315252  
QY 979 AGCGGCTCTATGAAGATGCGT---CGCAAGACCGCTTGACACACAGATATGATC 1035  
Db 315253 AGTGCAATTTATGAAGACGGAGTAAGACAAAGAAATAAAGAAACGAAGAAATATA 315312  
QY 1036 GACCACTGGAAGCACCATCTCGTGCAGAA 1065  
Db 315313 GAAAAATGGAATTAATAATTCGAAAAAA 315342

RESULT 4  
US-09-974-300-1692  
; Sequence 1692, Application US/09974300  
; Patent No. US20020146721A1  
; GENERAL INFORMATION:  
; APPLICANT: Berka, Randy M.  
; APPLICANT: Clausen, Ib Groth  
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene Expression  
; FILE REFERENCE: 10085.500-US  
; CURRENT APPLICATION NUMBER: US/09/974,300  
; PRIOR FILING DATE: 2001-10-05  
; PRIOR APPLICATION NUMBER: 09/680,598  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: 60/279,526  
; PRIOR FILING DATE: 2001-03-27  
; NUMBER OF SEQ ID NOS: 8481  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1692  
; LENGTH: 1083  
; TYPE: DNA  
; ORGANISM: Bacillus licheniformis  
US-09-974-300-1692

Query Match 26.2%; Score 293.4; DB 10; Length 1083;  
Best Local Similarity 56.6%; Pred. No. 3.4e-89;  
Matches 543; Conservative 0; Mismatches 416; Indels 0; Gaps 0;

QY 22 CAACGTAGAAAATCAACACGATATTTACGTTGGGAATGTGCGGATTTGGCGATGTGTCCTCC 81  
Db 21 CATCGTACAAAACGGCGTCCGTTAAAGTGGACCTTTAAACATAGGCGCAATAACGAA 80  
QY 82 ATCGCGTACAGTCCATGACCAATACGCTACACAGAGTCGAAGCAACGGTCAATCAA 141  
Db 81 GTCGTCAATTCAAAGCATGACAAACGAAACACATGACGTTGAAGCAACCGTCGCCGAA 140  
QY 142 ATCAAGGCGCTGGAACGCGTTGGCGTGATATCGTCCGTGTATCCGTACCGACGATGGAC 201  
Db 141 ATCAACAGACTCGGGAAGCAGAGATGTCAATCGTCCGCTGCCCTGTCTCTGATGAACGG 200  
QY 202 GCGGAGAAAGCGTTCAAATCATCAACACAGCAGTTAAAGTGGCGCTGGTGGCTGACATC 261  
Db 201 GCTGCGAGCCATTCACAGAGATCAAAAGCGATATCCATCCCTCTTGTCTGGATATT 260  
QY 262 CACTTCGATATCGCATTTGGCTGAAAGTAGCGGAATACGGCGTCGATTTGTCTGCGTATT 321  
Db 261 CATTTCAACTATAAATTTGGCATTTAAAGCGATCGAAGGGGAGCCGATATAAATCCGCATC 320  
QY 322 AACCTTGGCAATATCGGTAATGAAGCGGTATTTCGATGTTGTTGCTGCTGCGCGCAT 381  
Db 321 AATCCGGTAAACATCGCGCCCGCGAAAGGTTGAAGCGGTCTCAACGCGGAAGGAA 380  
QY 382 AAAAAATTCGATCCGATTTGCGTTAAAGCCGCGATCGCTGGAAAAAGATCTGCAAGAA 441  
Db 381 AAGGCGATTCGATCCGATTCGCGTCAATGACAGGCTCTCTGGAATAACGAATCCTTGAG 440  
QY 442 AGTATGGGAACCGACCGCGAGGCGTTGCTGGAATCTGCGCATGCGGTCAATGTTGATCAT 501  
Db 441 AAGTACGCGTATCCGACAGCAGCGCATGGTGGAAAGCGCCCTCGACCAACATTAAAT 500

QY 502 CTCGATCGCTGCACTTCGATCAGTTCAAAGTCAGCGTGAAGCGCTCTGACGTTCTCTC 561  
Db 501 CTTGAGGATCTCGATTTCCACGATATCATCGTCAGCATGAAGCGCTCTGATGTAACCTG 560  
QY 562 GCTGTTGAGTCTTATCGTTTGTGCAAAACAGATCAGCGGTTGATCTGCGGATC 621  
Db 561 GCGATTGAGGCATATGAAAAAGCGGCTAAAGCTTTTCGATTATCCGCTTCATTTAGGCATC 620  
QY 622 ACCGAAGCGGTGGTGGCGCGCAGCGGCGAGTAAATCCGCATTCGTTTATAGTCTGCTG 681  
Db 621 ACCGAATCGGAACCGCTGTTTGGCGGTAAGTAAAGCGCGCGGCTCTCGCGCGATC 680  
QY 682 CTGCTCTGAAGGCATCGCGCACACGCTGCGCGTATCGCTGCGCGCCGATCCGCTCGAAGAG 741  
Db 681 CTTTCAAAAGGAATCGGCAATACATTTGCGGATTTCTTTAAGCGCGACCCGCTCGAAGAA 740  
QY 742 ATCAAAGTCGTTTTCGATATTTTGAATTCGCTGCGTATCCGTTTCGCGAGGATCAACTTC 801  
Db 741 GTAAAAGTCGCCAGGAGCTCTCTGAAATCTTTTCGGGCTCGCTTCCAAATCGGCAACATTG 800  
QY 802 ATCGCTCGCCGACCTGTTTCGGGTGAGGAATTTGATTTATCGGTACGTTAACGCGCTG 861  
Db 801 ATTCTCTGCCGACTTTCGCGCGGATCGAAATCGATTGATCTGCAATTCGCAATGAATTC 860  
QY 862 GAGCAACGCTGGAAGATATCATCTCGATGCGAGCTTTTCGATTATCGGCTGCGTGGTG 921  
Db 861 GAAGATTACATCGGAAATCAAGCGCGGATCAAGTTGCGGCTCTCGGCTGCGGTC 920  
QY 922 AATGCCCAAGTGAGCGCTGTTTTCACATCGGCGTCAACGCGGCGCAACAGAAAG 980  
Db 921 AACGGTCCCGAGAGCGCGGAGCCGATATCCGATCGCGGCGCGCGGTAATG 979

RESULT 5  
US-10-156-761-2548  
; Sequence 2548, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 2548  
; LENGTH: 1155  
; TYPE: DNA  
; ORGANISM: Streptomyces avermitilis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1155)  
US-10-156-761-2548

Query Match 23.7%; Score 265; DB 14; Length 1155;  
Best Local Similarity 55.5%; Pred. No. 1.8e-79;  
Matches 533; Conservative 0; Mismatches 425; Indels 3; Gaps 1;

QY 43 ATTTACGTTGGAAATGTCGCGATTTCCCATGCTGCTCCCATCGCGTACAGTCCATGACC 102  
Db 70 ATCCAGGTGCGAACCCGTCGCGCGGAGACGACCGCTCTCGGTGCGATGAGC 129  
QY 103 AATACGCTTACACAGAGCTCGAAGCAACCGTCAATCAATCAAGCGCGCTGGAACGCGTT 162







Db 61 CCgATTGGCGATGGTGTCCATCGCGCTACAGTCCATGACCAATACGGCTACGACAGAC 120  
QY 121 GTCGAGCAGCGTCAATCAATCAAGCGCTGGAAGCGTTGGCGCTGATATCGTCCGT 180  
Db 121 GTCGAGCAGCGTCAATCAATCAAGCGCTGGAAGCGTTGGCGCTGATATCGTCCGT 180  
QY 181 GTATCCGTTACCGACGATGAGCGCGGAGAGCGTTCAAACTCATCAAAAGAGGTTAAAC 240  
Db 181 GTATCCGTTACCGACGATGAGCGCGGAGAGCGTTCAAACTCATCAAAAGAGGTTAAAC 240  
QY 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCAATTCGCGTGAAGTAGCGGAATAC 300  
Db 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCAATTCGCGTGAAGTAGCGGAATAC 300  
QY 301 GGCGTCGATTGTCTGGTATTAAACCTGGCAATATCGGTAAAGAGAGGTTATTCGCATG 360  
Db 301 GGCGTCGATTGTCTGGTATTAAACCTGGCAATATCGGTAAAGAGAGGTTATTCGCATG 360  
QY 361 GTGGTTGACTGTGCGCGCGATAAAAACATTCGATCCGTTATGGCGTTAAACGCGGATCG 420  
Db 361 GTGGTTGACTGTGCGCGCGATAAAAACATTCGATCCGTTATGGCGTTAAACGCGGATCG 420  
QY 421 CTGGAAAAAGATCTGCAAGAAAAAGTATGCGAAACCGACCGCGAGCGGTTGCTGGAATCT 480  
Db 421 CTGGAAAAAGATCTGCAAGAAAAAGTATGCGAAACCGACCGCGAGCGGTTGCTGGAATCT 480  
QY 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTCAAGTACAGCGTG 540  
Db 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTCAAGTACAGCGTG 540  
QY 541 AAAGCGCTCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTGTCGCAAAACAGATCGAT 600  
Db 541 AAAGCGCTCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTGTCGCAAAACAGATCGAT 600  
QY 601 CAGCGCTTGATCTGGGATCACGAGCGGTTGTCGCGACAGCGGGCAGTAAATCC 660  
Db 601 CAGCGCTTGATCTGGGATCACGAGCGGTTGTCGCGACAGCGGGCAGTAAATCC 660  
QY 661 GCATTGTTTAGTCTGCTGCTGCTGAGGATCGGACAGCGCTGCGCGTATCGCTG 720  
Db 661 GCATTGTTTAGTCTGCTGCTGCTGAGGATCGGACAGCGCTGCGCGTATCGCTG 720  
QY 721 GCGGCGATCCGCTCGAAGAGATCAAGTCCGTTTCGATATTTGAAATCGCTGCGTATC 780  
Db 721 GCGGCGATCCGCTCGAAGAGATCAAGTCCGTTTCGATATTTGAAATCGCTGCGTATC 780  
QY 781 CGTTCGCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAATTTGATGT 840  
Db 781 CGTTCGCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAATTTGATGT 840  
QY 841 ATCGGTACGGTTAACGGCTGGAGCAACGCTGGAAGATATCATCTCGATGGACGTT 900  
Db 841 ATCGGTACGGTTAACGGCTGGAGCAACGCTGGAAGATATCATCTCGATGGACGTT 900  
QY 901 TCGATTATCGGCTCGCTGGTGAATGCGCAGGTGAGCGCTGTTCTACACTCGGCGTC 960  
Db 901 TCGATTATCGGCTCGCTGGTGAATGCGCAGGTGAGCGCTGTTCTACACTCGGCGTC 960  
QY 961 ACCGCGGCAACAGAAAAAGCGCCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020  
Db 961 ACCGCGGCAACAGAAAAAGCGCCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020  
QY 1021 AACACGATATGATCCAGCAGCTGGAAGCAGCATTCGTCGGAAGCCAGCTCAGCTGGAC 1080  
Db 1021 AACACGATATGATCCAGCAGCTGGAAGCAGCATTCGTCGGAAGCCAGCTCAGCTGGAC 1080  
QY 1081 GAAGCGGCTCGAATTGACGTTACAGCAGGTTGAAAAATAA 1119  
Db 1081 GAAGCGGCTCGAATTGACGTTACAGCAGGTTGAAAAATAA 1119

; Sequence 1, Application US/10329960  
; Publication No. US2003009277A1  
; GENERAL INFORMATION:  
; APPLICANT: Fleischmann et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm  
; FILE REFERENCE: PB186P1  
; CURRENT APPLICATION NUMBER: US/10/329,960  
; CURRENT FILING DATE: 2003-01-02  
; PRIOR APPLICATION NUMBER: US 09/643,990  
; PRIOR FILING DATE: 2000-08-23  
; PRIOR APPLICATION NUMBER: US 08/487,429  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/426,787  
; PRIOR FILING DATE: 1995-04-21  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 1  
; LENGTH: 1830121  
; TYPE: DNA  
; ORGANISM: Haemophilus influenzae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (4747)..(4747)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (9921)..(9921)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (10150)..(10150)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (29298)..(29298)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36543)..(36543)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36551)..(36551)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (36636)..(36636)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (40808)..(40810)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (44416)..(44416)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (44905)..(44905)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (44975)..(44975)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (45593)..(45593)  
; OTHER INFORMATION: n equals a, t, g or c  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (45732)..(45732)  
; OTHER INFORMATION: n equals a, t, g or c

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 15:08:02 ; Search time 412 Seconds  
(without alignments)  
8876.507 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataacaggctccaat.....ttcagcaggtgaaaaataa 1119

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 15: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 16: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 17: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1119	100.0	1119	9	US-09-921-992-3
2	622	55.6	1830121	14	US-10-329-960-1
3	403.6	36.1	640681	10	US-09-790-988-1
4	293.4	26.2	1083	10	US-09-974-300-1692
5	265	23.7	1155	14	US-10-156-761-2548
6	265	23.7	9025608	14	US-10-156-761-1
7	260.6	23.3	1155	14	US-10-156-761-1636
8	243.8	21.8	1164	10	US-09-712-363-1110
9	239.4	21.4	1134	10	US-09-738-626-2205
10	239.4	21.4	3309400	10	US-09-738-626-1
11	211.8	18.9	1180	10	US-09-881-752A-195
12	145.6	13.0	706	14	US-10-066-543-1184
13	70.6	6.3	584	9	US-09-921-992-21
14	68.4	6.1	596	9	US-09-921-992-23
15	68.4	6.1	670	9	US-09-921-992-22
16	56.8	5.1	348	10	US-09-974-300-6102

17	52.6	4.7	293	9	US-09-921-992-25	Sequence 25, Appl
18	52.6	4.7	456	9	US-09-921-992-26	Sequence 26, Appl
19	50	4.5	403	9	US-09-921-992-24	Sequence 24, Appl
20	48.8	4.4	564	9	US-09-921-992-35	Sequence 35, Appl
21	48.4	4.3	379	9	US-09-921-992-16	Sequence 16, Appl
22	48.4	4.3	528	9	US-09-921-992-15	Sequence 15, Appl
23	48.4	4.3	2520	9	US-09-921-992-1	Sequence 1, Appl
24	48.2	4.3	353	9	US-09-921-992-39	Sequence 39, Appl
25	48.2	4.3	464	9	US-09-921-992-33	Sequence 33, Appl
26	47.2	4.2	601	9	US-09-921-992-11	Sequence 11, Appl
27	45.4	4.1	613	9	US-09-921-992-32	Sequence 32, Appl
28	45.2	4.0	432	9	US-09-921-992-14	Sequence 14, Appl
29	45	4.0	938	9	US-09-921-992-13	Sequence 13, Appl
30	43.6	3.9	443	9	US-09-921-992-12	Sequence 12, Appl
31	43.4	3.9	291	9	US-09-921-992-43	Sequence 43, Appl
32	41.8	3.7	577	9	US-09-921-992-40	Sequence 40, Appl
33	41	3.7	422	9	US-09-921-992-28	Sequence 28, Appl
34	41	3.7	430	9	US-09-921-992-29	Sequence 29, Appl
35	40.2	3.6	33675	9	US-09-921-992-2	Sequence 2, Appl
36	39.6	3.5	671	14	US-10-184-644-346	Sequence 346, App
37	39.6	3.5	671	14	US-10-184-634-346	Sequence 346, App
38	38.8	3.5	388	9	US-09-921-992-44	Sequence 44, Appl
39	37.2	3.3	511	9	US-09-921-992-36	Sequence 36, Appl
40	37.2	3.3	705	9	US-09-921-992-34	Sequence 34, Appl
41	36.6	3.3	395	9	US-09-921-992-18	Sequence 18, Appl
42	36.6	3.3	1714	14	US-10-242-056-60	Sequence 60, Appl
43	36.4	3.3	16596	14	US-10-156-761-930	Sequence 930, App
44	36.4	3.3	135638	12	US-10-314-657-1	Sequence 1, Appl
45	35.6	3.2	211	9	US-09-921-992-45	Sequence 45, Appl

#### ALIGNMENTS

#### RESULT 1

US-09-921-992-3  
; Sequence 3, Application US/09921992  
; Patent No. US20020069426A1  
; GENERAL INFORMATION:  
; APPLICANT: Boronat, Albert;  
; APPLICANT: Campos, Narciso;  
; APPLICANT: Rodriguez-Concepcion, Manuel;  
; APPLICANT: Rohmer, Michel;  
; APPLICANT: Seeman, Myriam;  
; APPLICANT: Valentin, Henry E.;  
; APPLICANT: Venkatesh, Tyamagondlu V.;  
; APPLICANT: Venkatamesh, Mylavaram  
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes  
; FILE REFERENCE: 16516.107/35-21(51897)US  
; CURRENT APPLICATION NUMBER: US/09/921,992  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US 60/223,483  
; PRIOR FILING DATE: 2000-08-07  
; NUMBER OF SEQ ID NOS: 85  
; SEQ ID NO 3  
; LENGTH: 1119  
; TYPE: DNA  
; ORGANISM: Escherichia coli  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1119)  
US-09-921-992-3

Query Match 100.0%; Score 1119; DB 9; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCATACCCAGGCTCCCAATTCACGTCAGAAATCAACACGTATTACCTGGGAATG	60
DB	1	ATGCATACCCAGGCTCCCAATTCACGTCAGAAATCAACACGTATTACCTGGGAATG	60
QY	61	CCGATTGGCGATGGTGTCTCCCAATCCCGTACAGTCCATCCCAATACGCTGACGAC	120

**THIS PAGE BLANK (USPTO)**

Db 541 GCTCCGGCACGGTGA 555

Search completed: November 23, 2003, 15:55:33  
Job time : 96 secs

	Query Match	35.9%	Score 401.4	DB 4	Length 1137
;	TYPE: DNA				
;	ORGANISM: Acinetobacter baumannii				
US-09-328-352-3780	Best Local Similarity 60.8%				
	Matches 654; Conservative 0; Mismatches 421; Indels 0; Gaps 0				
Qy	11	AGGCTCAAATTC	AAAGTAA	CAACACGCTATT	TACGTTTGGGAATGTCGATGGCG 70
Db	29	AGAACCAATTTAA	ACGTCGACCA	ACGCTAAATCCG	TGGTTCGATGTCGCTG 88
Qy	71	ATGCTGCTCCCAT	CGCGTACAGTCC	ATGACCAATACCG	TACGACAGCTCGAAGCA 130
Db	89	CGCATGACCTTAT	TAGTGTGCAAA	GTATGACAAATAC	CGAACTTTCGATGTGTGACGCA 148
Qy	131	CGGTCAATCAAT	CAAGCGCTGGA	ACCGTTGGCGCTG	ATATCGTCCGTGTATCCGTAC 190
Db	149	CTGTGGCTCAG	ATTGAGCGTTGT	TGATGACAGGTG	CAGATATATGCGTGTTCAGTGC 208
Qy	191	CGCATGGACGCG	CAGAAAGCTTT	CAAACTCATCAAA	CAGCAGGTAAACGTGCGCGTGG 250
Db	209	CTTCTATGGA	AGCTGCTGAAG	CCCTTTTGGTGC	AAATCCGTAAGCGTGTTCAGTTCATTAG 268
Qy	251	TGGCTGACATCC	ACTTCGACTATCG	CATTGCGCTGAA	AGTAGCGGAATACGCGCTCGATT 310
Db	269	TAGCTGATATCC	ATTTGACCATAGA	ATTGCTTTGGCAG	TTTGCAGATTATGGTGCGGACT 328
Qy	311	GTCTGCGTATT	AACCCCTGGCA	ATATCCGTTAAT	GAAAGCGTATTCGTCATGGTGGTTGACT 370
Db	329	GCTTTCGTTAT	TAAACCGGGCA	ATATCCGTTT	CAGACCAAGAGTTCGTGAAGTTGTGGCTG 388
Qy	371	GTGCGCGCAT	TAATAAACAAT	CCGATCCGTTAT	TGGCGTTAACCGCCGATTCGCTGGAATAAG 430
Db	389	CGGCACGTATC	ACGGTATTTCT	TATGCGTATTTG	CGGTGAATCGCGGTCTCTAGAAAAAG 448
Qy	431	ATCTGCAAGAA	AGATATGGCGAA	CCGACGCGCAGG	CGTTCGTGAAATTCGCCATCGCTC 490
Db	449	ATTTACAGAAA	AAATATATGGCAG	CCCTACAGGCGAG	CACTTCTTGAGTCAGCTTTACGCTC 508
Qy	491	ATGTTGATCAT	CTCGATCGCCTG	AACTTCGATCAG	TTCAAAGTCAGCGTGAAGCGTCTG 550
Db	509	ATATTGATATT	TTAGACCGTCTT	TGACTTCATGAG	TTTAAAGTCAGTGTAAAGCATCAA 568
Qy	551	ACGCTCTCTCG	CTGTGAGTCTTAT	CGTTTCTGCGCAA	ACAGATCGATCAGCGTTCG 610
Db	569	ATGTGTTTTTAA	CAACATGATGCT	TATCGTTTACTCT	CTCAACAAATTGATAATCCATTAC 628
Qy	611	ATCTGGGGATC	ACCGAGCCGCTG	TGCGCGCAGCGG	GCAGTAAATCCGCCATTGGTT 670
Db	629	ACCTTGGAGT	GACTGAAGCTGG	ATTATCCGTACAG	GTACTGTGAATCAGCGATTGCTC 688
Qy	671	TAGTCTGCTCTG	CTCTGAAGGCAT	CGGCGACACGCT	CGCGTATCGTGGCGGCGCATC 730
Db	689	TTGCTGGGTATT	GATGAAGGCATT	TGGCGATACATG	CGGTATTTTCGCTGCTGTGAAC 748
Qy	731	CGGTGCAAGAG	ATCAAAAGTCGG	TTTCGATATTTTG	AAATTCGCTGCGTATCCGTTTCGCGAG 790
Db	749	CTGAAGATGAA	ATCAAGATCGGT	TTTGATATCTTAA	AAATCGTTGCGCTACGTTCTTAACG 808
Qy	791	GGATCAACTTCA	TGCGCTGCCGAC	CTGTTCCGCTCAG	GAATTCATGTTATCGGTACGG 850
Db	809	GTAATCAACTT	TATGCTTGTCCA	AGTGTTCGCGCA	AGAAATTTAACGTGATTCAGGTGA 868
Qy	851	TTAAACGGCTG	GGAGCAACGCTG	GGAAGATATCAT	CACCTCCGATGGAACGTTTCGATTATCG 910
Db	869	TGCAGGCTTTA	GAGAGCGTTTAG	AAGATATCCGTAC	CAACATGACGCTTCGTTTCGTTATTG 928
Qy	911	GCTCGGTGGTA	ATGGCCAGGTG	AGGCGCTGTTTCT	ACATCGCGGTCAACCGCGGCA 970
Db	929	GCTGTAAAGT	TAATATGGCCAG	GTGAAGCAAAAG	ACAGATATATCGGGTTGTTCGGGCTG 988

QY 661 GCCATTGGTTTAGTCTGCTGCTCTGAAGGATCGGCGACACGCTGCGCGTATCGCTG 720  
|||  
Db 1467 GCGTGGGGCTGGGCATGCTCTGGCGGAGGAATCGGCGACACCATCCGGAATTCCTTG 1408  
|||  
QY 721 GCGGCGATCCGGTCGAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGTCGGTATC 780  
|||  
Db 1407 GCTGCCGATCCGGTCGAGGAGATCAAGTCGGTTTCGACATCCTCAAGTCCCTGCACCTG 1348  
|||  
QY 781 CGTTCCGAGGGATCAACTTCATCGCTGCCCGACCTGTTTCGCGTCAGGAATTTGATGT 840  
|||  
Db 1347 CGTCCCGTGGCATCAACTTCATCGCTGCCCGACCTGTTTCGCGGAGAACTTCGACGTG 1288  
|||  
QY 841 ATCGGTACGGTTAACCGCTGGAGCAACGCTCGAAGATATCATCTCCGATGGACGTT 900  
|||  
Db 1287 GTGAAGACCATGAACGAGCTGGAGGGCGCTGGAGGACCTGCTGGTCGATGGACGTG 1228  
|||  
QY 901 TCATTTATCGGTCGGTGGTGAATGCGCCAGTGGAGCGCTGTTTCTACACTCGCGCTC 960  
|||  
Db 1227 GCGTGTATCGTTGCGCTCGTCAACGGTCCGCGGGAAGCCAAAGGAGGCCATGTCCGCTC 1168  
|||  
QY 961 ACCGGCGGCAACGAAGACGGCCCTCTATGAAGATGGCGTCGCGAAGACCGTCTGGAC 1020  
|||  
Db 1167 ACCGGCGGCACTCCGAA---CCTGGTGTATCGACGGCAAGCCGTGCGAAGAACTGACC 1111  
|||  
QY 1021 AACCAAGATATGATCGACCAAGCTGGAAGCAACGATTCGTGCGAAAGCCAGTCAAGTGGAC 1080  
|||  
Db 1110 AAGCAACACTGTGTGACGAGCTGGAAACGGCTGATCGGCGAAGGGCGGCGAAGAGGCC 1051  
|||  
QY 1081 GAAGC 1085  
|||  
Db 1050 GAGGC 1046  
|||

## RESULT 13

US-09-252-991A-6339  
; Sequence 6339, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6339  
; LENGTH: 2088  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6339

Query Match 45.8%; Score 512; DB 4; Length 2088;  
Best Local Similarity 68.0%; Pred. No. 1.8e-151;

Matches 729; Conservative 0; Mismatches 340; Indels 3; Gaps 1;

QY 14 CTCCAATTCAACGTAGAAAATCAACACGTATTTACCGTGGGAATGTCGCGATTGGCGATG 73  
|||  
Db 6 CTCGGATCATTCGCCCAAGTCTCGGAAATCTGGTTCGGCAACGTCCCGTGGCGCGCG 65  
|||  
QY 74 GTGCTCCCATCCGGTACAGTCCATGACCAATACCGGTACGACAGACGTGCGAAGCAACGG 133  
|||  
Db 66 ACGCGCGGATCCGGTGCAGAGCATGACCAACACCGAGACCTGCGACGTCGCTGCCACCG 125  
|||  
QY 134 TCAATCAATCAAGGCGCTGGAACGGTTGGCGGTATATCGTCCGTGTATCGGTACCGA 193  
|||  
Db 126 TCGCGAGATCCGCCCTCGGAAGATGCGCGCGGAGACATCGTGGCGGTTTCGTCGCCG 185  
|||  
QY 194 CGATGGACCGGCAAGAGCGTTCAAACTCATCAACACGAGGTTAACGTGCGCGTGGTG 253  
|||

Db 186 ACATGAGCGCCGCGAGGCAATTCGCAAGATCAAGCAGCAGGTCAACGTGCGCTGGTCG 245  
|||  
QY 254 CTGACATTCACATTCGACTATCGCATTCGCTGAAAGTAGGGNATACGGCGTGCATGTGTC 313  
|||  
Db 246 CCGACATTCACATTCGACTATCGCATTCGCTGCGCGTTCGCGAGTGGGAGTGCATGCC 305  
|||  
QY 314 TCGGTATTAACTCGTCAATATCGGTAAATGAAGAGCGTATTCGATGGTGGTTGACTGTG 373  
|||  
Db 306 TGGGCATCAATCGGGCAACATCGGTGCGGAGGACCGGGTCAAGGCGGTGGTTCGATGCCG 365  
|||  
QY 374 CCGCGGATAAAAACAATTCGATTCGATTCGCGTAAACCGCGGATCGCTGGAAAAAGATC 433  
|||  
Db 366 CCGCGAGCGCAACATCCCGATCCGTTATCGCGTCAATCGCGTTCGCTGGAAAAAGACC 425  
|||  
QY 434 TGCAGAAAGATATGCGGAACCGACCGCGAGCGGTGCTGGAATCTGCCATCGCTCATG 493  
|||  
Db 426 TGCAGAAAGATATGCGGAACCGACCGCGAGCGGTGCTGGAATCTGCCATCGCGCACG 485  
|||  
QY 494 TTGATCATCTCGATCGCTGAATTCGATCAGTTCAGATTCAGGTGAAAGCGTCTGACG 553  
|||  
Db 486 TGCATCATCTCGAAGCTGGACTTCCAGAACTTCAAGTTCAGGTTCAGGCGCTCCGACG 545  
|||  
QY 554 TCTTCCTCGCTGTGAGTCTTATCGTTTCTGCTGGCAAAACAGATCGATCAGCCGTTGCATC 613  
|||  
Db 546 TCTTCATGCGCTGCGCGCTATCGCTGCTGGCAGGAGATCGAGAGCCCTCGAC 605  
|||  
QY 614 TGGGATACCAAGCGGTGTCGCGAGCGGGGAGTAAATCCGCGATTTGGTTTAG 673  
|||  
Db 606 TGGGCATCACCGAGGCGCGCGCTGCGCTCGGACCGTGAAGTCCGCGTGGGCTGG 665  
|||  
QY 674 GTCTGCTGTCTGAAGGCATCGGCGACACGCTGCGGTATCGCTGGCGCGCGATCCGG 733  
|||  
Db 666 GCATGCTCTGCGCGAGGAATCGGCGACCATCCGATTCGCGTGGCTGCCGATCCGG 725  
|||  
QY 734 TCAGAGAGATCAAGTTCGTTTCGATATTTTGAATTCGCTGGGTATCCGTTTCGCGAGGGA 793  
|||  
Db 726 TCGAGGAGATCAAGTTCGTTTCGATATTCGATATTCGATATTCGATATTCGATATTCGAT 785  
|||  
QY 794 TCAACTTCATCCCTGCGCGAGCTGTTGCGGTGAGAACTTCGACGTTGTAAGGTTA 853  
|||  
Db 786 TCAACTTCATCCCTGCGCGAGCTGTTGCGGTGAGAACTTCGACGTTGTAAGGTTA 845  
|||  
QY 854 ACGCGTGGAGCAACGCTGGAAGATATCATCACTCCGATGACGTTTCGATATTCGCT 913  
|||  
Db 846 ACGAGCTGGAAGGCGCTTGGAGGACCTGCTGGTCCGATGAGCGTGGCGGCTGATCGGTT 905  
|||  
QY 914 GCGTGTGAATGGCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTCAACGCGGCAACA 973  
|||  
Db 906 GCGTGTCAACGCTCGGCGGAAGCCAGGAGGCCCATGTCCGCTCACCGCGGCACTC 965  
|||  
QY 974 AGAAAAAGGCGCTTATGAAGATGCGGTGCGCAAGACCGTCTGGACAAACATATGA 1033  
|||  
Db 966 CGAA---CCTGGTGTATATCGACGCGCAAGCCGCTCGCAGAAACTGACCAACGCAACTCG 1022  
|||  
QY 1034 TCGACCAAGTGAAGCAACGATTCGTTGCGAAAGCCAGTTCAGCTGGACGAGC 1085  
|||  
Db 1023 TGGACAGCTGAACGCGTGTATCCGCCAGAGGCGCGCGAGAAAGCCGAGGC 1074  
|||

## RESULT 14

US-09-328-3780  
; Sequence 3780, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Berton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 3780

Qy	1	ATGATTAAC	CGAGGTCCAA	TTTCAAGCTAG	AAAAATCA	ACAGTATTT	TACGTTGG	AAATGTG	60
Db	535	ATT	CATAGCG	TCTTCCGAT	CAITTCGCG	CAAGTCTCGG	AAAAATCTGG	TTCGGCAAGTC	594
Qy	61	CCGATT	CGCGATG	TGCTCCCAT	CGCGGTAC	AGTCCAT	GACCAAT	ACGCGTAC	120
Db	595	CCG	TGGGCGG	AGCGCCGAT	CGCCGTG	CAGAGCAT	GACCAAC	CCGAGACT	654
Qy	121	GT	CGAAGCA	ACGGTCAAT	CAAAATCA	AGGCGCTG	GAAACGGT	TGCGTGTAT	180
Db	655	GT	CGCTGCC	ACCGTTCG	CGAGATCG	CCGCCCTG	GAAGATCG	CGCGGACAT	714
Qy	181	GTAT	CCGTA	CGGACGAT	GGAGCGG	CGAGAGCG	TTCAAACT	CAACACAG	240
Db	715	GT	TTCCGTC	CCCGACAT	GGACCG	CGCGAGGCA	TTTCGGCA	AGATCA	774
Qy	241	GTG	CCGCTG	TGGGTG	ATCCACT	TCGACTAT	CGCATTC	CGCTGAA	300
Db	775	GTG	CCGCTG	TGCGCG	ATCCATTC	GATCGCAT	TGCGCCT	TGCGGCT	834
Qy	301	GGCG	TGATTGT	CTGCGT	ATTAAAC	CCCTGG	CAATAT	CGTAAAG	360
Db	835	GGAG	TGACTG	CTGCGCAT	CAATCC	GGGCAACAT	CGTTCG	GAGGAC	894
Qy	361	GTG	TTGACTG	TGCGCGG	ATAAAAA	CATTC	CGATTC	CGGTAA	420
Db	895	GTG	TGATG	CGCGCG	CGAGCAAC	ATCC	CGATTC	CGGTCA	954
Qy	421	CTG	GAANA	AGATCTG	CAAGAA	AGTATG	GGGAAC	CGGAC	480
Db	955	CTG	GAANA	AGACCTG	CAAGAA	ATATCG	GGGAAC	CCGAC	1014
Qy	481	GCA	TGCGT	CATGTT	GTATCAT	CTCGAT	CGCTG	AACCTTC	540
Db	1015	GCA	TGCGCAC	GTGATCAT	CTCGAC	NAAGTCT	CCAGACTT	CAAA	1074
Qy	541	AA	AGCTGT	GACGTCTT	CTGCTG	TGTTG	AGTCTT	ATCGTT	600
Db	1075	AAG	SCCTCC	GACGCTT	CATGG	CGCTG	CGCGCTAT	CGCTG	1134
Qy	601	CAG	CGTTG	ATCTGGG	ATCAC	GAAGCCG	TGTTGG	CGACG	660
Db	1135	CAG	CCCTG	CACTGGG	ATCAC	GAAGCCG	CGCGCTG	CGCGA	1194
Qy	661	GCA	ATTG	TTTAG	CTGCTG	CTGTGA	AGGCAT	CGCGAC	720
Db	1195	GCG	TGGGCT	TGGCATG	CTCTG	CGCGAG	GAATCG	CGGAC	1254
Qy	721	CG	GCGCAT	CCGTC	GAGAGAT	CAAA	GTG	TTTCGAT	780
Db	1255	GCT	GCCGAT	CCGTC	GAGGAGAT	CAAG	GTG	TTTCGAT	1314
Qy	781	CGT	TCGAG	GGGATCA	ACTTCAT	CGCTG	CCCCAC	CTTTCG	840
Db	1315	CGT	CCCGT	GGCATCA	ACTTCAT	CGCTG	CCCCAG	CTTTCG	1374
Qy	841	AT	CGGTAC	GGTTAA	CGCGCTG	GACCAAC	CGCTG	GAAGAT	900
Db	1375	GT	GAAGAC	CCATGA	ACGAGCTG	GAAGGG	CGCTG	GAGAC	1434
Qy	901	TC	GATTA	TGCGCT	CGTGGT	GAATG	GCCCCAG	GTGAG	960
Db	1435	GCC	GTGAT	CGGTTG	CGTTCG	TCAAC	CGGTCC	GGCGA	1494
Qy	961	AC	CGGCGG	CAACA	GAAGGG	CGCTTAT	GAAATG	CGGTG	1020
Db	1495	AC	CGGCGG	CACTCC	GAA---	CGTGT	GATAT	CGACG	1551
Qy	1021	AACA	ACGAT	TGATG	TCGAC	AGCTG	GAAGCA	CGCAT	1080
Db	1552	AACA	CAAC	CTGTG	TGAGAG	CTGGA	ACGGCTG	ATCC	1611
Qy	1081	GA	AGC	1085					

```

Db          1612 GAGGC 1616
|||||
RESULT 12
US-09-252-991A-6131/c
; Sequence 6131, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6131
; LENGTH: 2202
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6131

Query Match          46.1%; Score 515.4; DB 4; Length 2202;
Best Local Similarity 67.8%; Pred. No. 1.6e-152;
Matches 736; Conservative 0; Mismatches 346; Indels 3; Gaps 1;

QY 1 ATGCATTAACGAGGCTCCAAATTCACACGTAGAAAAATCAACACGTAATTTACGTTGGGAATGTG 60
DB 2127 ATTCATAGCGCTTCTCCGATCATTCGCCGCAAGTCTCGGAAAAATCTGGGTCCGCCAACGTC 2068

QY 61 CCGNTTGGCGATGGTGTCTCCATCGCGGTACGTTCATGACCAATACCGCTACGACAGAC 120
DB 2067 CCGGTGGCGGCGACGCGCGCATCGCGGTGCAGACATGACCAACCCGAGACTCTCGCAC 2008

QY 121 GTCGAAGCAACGGTCAATCAAATCAAGGCGCTCGAAGCGCTTGGCGCTGATATCGTCCGT 180
DB 2007 GTCGCTGCACCTGCGCGAGATTCGCCGCTTGGAGATGCCGGCGGACATCTGTCGG 1948

QY 181 GTATCCGTACCGAGATGGACGCGGCAGAGCGTTCAAACCTCATCAAAACAGCAGGTTAAC 240
DB 1947 GTTTCCGCTCCGACATGGACGCGCGCGAGGCATTTCGGCAAGATCAAGCAGCAGGTC AAC 1888

QY 241 GTGCGCGTGTGGTGNACATCCACTTCGACTATCGCATTCGCGCTGAAAGTAGCGGAATAC 300
DB 1887 GTGCGCGTGTGCGCGACATCCACTTCGACTATCGCATTCGCCCTGCGCGCTGCGCGAGCTG 1828

QY 301 GGCCTGCATTGTCTGCGTATTAACCCCTGGCAATATCGGTAATCAAGAGCGTATTTCGCATG 360
DB 1827 GGAGTGACTGCTGCGCATCAATCCGGGCACAACTCGTTCGAGGACCCGGTCAAGGCC 1768

QY 361 GTGCTTGACTGTGCGCGCGATAAAAAATTCGATCCGATCCGTTATGCGGTTAAACGCCGATCG 420
DB 1767 GTGCTGCATGCCGCGCGAGCGCAACATCCCGATCCGATCCGTTATCGCGCTCAATGCCGTTTCG 1708

QY 421 CTGGAAAAAGATCTGCNAGAAAAAGTATGGCGAACCCGACCCGAGCGGTTCGTTGGAACTCT 480
DB 1707 CTGGAAAAAGACCTGCNAGAAAAATACGGCGAACCCGACCCCGAAGCCCTGCTCGGAATCG 1648

QY 481 GCATGCGTTCATGTGTGATCATCTCGATTCGCTGAACTTCGATCAGTTTCAAAGTCACGCGTG 540
DB 1647 GCCATGCGCCAGCTCGATCATCTCGAACGCTGCACTTCCAGAACTTCAAGTCAAGCGTC 1588

QY 541 AAAGCGTCTGACGCTTTCCTCGCTGTTGAGTCTTATCTGTTTGTGGCAAAACAGATCGAT 600
DB 1587 AAGGCTCCGACGCTTTCATGGCCGTCGCGCCTATCCGCTGCTGSCCGAGCATCGAG 1528

QY 601 CAGCCGTTGCATCTGGGGATCAACCGAAGCCGCTGTTGGTGGCGCAGCGGGCAGTTAAATCC 660
DB 1527 CAGCCCTGCACCTGGGCATCACGAGGCGCGCGGCTGCGCTCCGGCACGCTGAAGTCG 1468

```

Db 166 TTGCTTTGCCAACCTGTTACGCCAAGAAATTTGATGTGATTGTCACGGTAAATGCTTTGG 107  
QY 863 AGCAAGCCCTGGAAGATATCATCTCCGATGACGTTTCGATTATTCGGCTGGTGTGA 922  
Db 106 AGCAGCGCTCGAAGATATATACCGCGATGATGTCTATATTGTTGTGTAGTGA 47  
QY 923 ATGGCCAGGTGAGGCGCTGGTTTCTACACTCGCGCTCACCGCGC 967  
Db 46 ATGGCCCGGTGAAGCCGAGGTTTCTACTTTAGTGTGGCTGGCG 2

RESULT 10  
US-09-170-187-10/c  
; Sequence 10, Application US/09170187  
; Patent No. 6381745  
; GENERAL INFORMATION:  
; APPLICANT: Racher, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medien & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/170,187  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/827,190  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; TYPE: nucleic acid  
; LENGTH: 886 base pairs  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-09-170-187-10

Query Match 47.8%; Score 534.6; DB 4; Length 886;  
Best Local Similarity 75.3%; Pred. No. 8.2e-159;  
Matches 666; Conservative 0; Mismatches 219; Indels 0; Gaps 0;  
QY 83 TGGCGGTACAGTCCATGACCAATACCGGTACGACAGCGTGAAGCAACGGTCAATCAA 142  
Db 886 TTGCTGTCCAATCTATGACCAATACCGGTACGACAGCGATGTTAGCGCACTGTGCGGCAA 827  
QY 143 TCAAGCGCTGGAACCGGTGGCGGTGATATCGTCCGTATCCGTACCGACGATGGAGC 202  
Db 826 TCCAATCACTTGAAGCGGTGATGATATCGTCCGCGTGTCTGTTCTACGATGGATG 767  
QY 203 CGGCAAGCGTTCAAACTCATCAACAGCAGGTAAACGTGCGCTGGTGGCTGACATCC 262  
Db 766 CAGCAGAGCCCTTAATTAATTAACAGCGGTGAATGTCATTTGGTTGGGATATTC 707  
QY 263 ACTTCGACTATCGCAITGGCGCTGAAAGTAGCGGAATACGCGGTTCGATTGCTCGCGTATTA 322

Db 706 ACTTTACTACCGTATCGCGATGAAAGTGGCTGAATATGGTTGACTCCCTACGAATTA 647  
QY 323 ACCCTGGCAATATCGTAAATGAAGAGCGTATTTCGATGTTGTTGACTCTGCGCGGATA 382  
Db 646 ACCAGGTATATCGCAGTGAAGAGCGTATTTCGCCAAGTCTGTTGATGTCTGCTCATC 587  
QY 383 AAAACATTCCGATCCGTATTGGCGTTAAACGCCGATCGCTGGAAAAAGATCTGCAAGAAA 442  
Db 586 ACAACATTCTATCCGTATAGGGTCAATGGCGGGTCACTGGAAAAAGATATCCAGAAA 527  
QY 443 AGTATGGGAACCGACCGCGAGGCTTCTCGAATCTGCCATGCGTCAATGTTGATCATC 502  
Db 526 AATACGGTGAAGCAACACCTGAAGCATTTGGTTGAATCAGCAATGCGACATGTTGATATCT 467  
QY 503 TCGATCGCTGAACCTCGATCAGTTTCAAGTTCAGGTGAGGCTGAAAGGCTGACGCTTCTCTCG 562  
Db 466 TGGACAGGCTGAATTTTCGATCAGTTTCAAGGTGAGTGTAAAGCGTGGGATGCTTTCTTG 407  
QY 563 CTGTTGAGTCTTATCGTTTGTCTGGCAAAACAGATCGATCAGCGGTTGCAATCTGGGGATCA 622  
Db 406 CCGTGGCTCTTATCGTTTATTTGGCGCAAAAATTTGATCAACCACTTCACCTCGGTATTA 347  
QY 623 CCGAAGCGGTGGTGGCGGACGCGGAGTAAATTCGCCATTTGGTTAGTCTGCTGC 682  
Db 346 CAGAAAGCGGTGGGCTCGTTCTGTTTCAAGTGAATCAGCAATTTGTTGTTATGTTGT 287  
QY 683 TGTCTGAAGGCATCGCGACACGCTGCGGTTATCGCTGCGGCGCGATCCGTCGAGAGA 742  
Db 286 TGGCTGAAGGTATCGCGGATACGTTACGTTATCTACTCGCGCAGATCCTGTTGAGGAAG 227  
QY 743 TCAAGTGGTTCGATATTTTGAATCGTCGCTATCGTTTCGGGAGGATCAACTTCA 802  
Db 226 TGAAGTGGTTCGATATTTTGAATCGTTACGATCGCTCAGTGGCATCACTTTA 167  
QY 803 TCGCTGCCGACCTGTTGGCTCAGGAATTTGATGTTATTCGGTACGGTTAAACGGCTCG 862  
Db 166 TTGCTTGGCCAACTGTTTCAAGCAAGAAATTTGATGTTGATTGTTGTTGTTGTTG 107  
QY 863 AGCAAGCGCTGGAAGATATCATCTCCGATGAGGTTTCGATTATCGGCTCGTGGTGA 922  
Db 106 AGCAGCGCTCGAAGATATTTATCAGCGCGATGGATGTTCTTATTATTGTTGTTAGTGA 47  
QY 923 ATGGCCCGGTGAGGCGCTGGTTTCTACACTCGCGGTCAACCGCGC 967  
Db 46 ATGGCCCGGTGAGCCGAGGTTTCTACTTTAGTGTGGCTGGCG 2

RESULT 11  
US-09-252-991A-6420  
; Sequence 6420, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 6420  
; LENGTH: 1644  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-6420

Query Match 46.1%; Score 515.4; DB 4; Length 1644;  
Best Local Similarity 67.8%; Pred. No. 1.3e-152;  
Matches 736; Conservative 0; Mismatches 346; Indels 3; Gaps 1;



Db 304 GTATCAATCCTGGCAACATTTGGTCGTGAAGATCGCTCCGCTGTTGATTTGGCGC 363  
QY 377 GCGATAAACAATCCGATCCGATTCGTTGGCGTTAAACCGCGATCGCTGGAAAAGATCTGC 436  
Db 364 GAGACAAAATATTCGATTCGTTGGTGTAAATCAGGCTCTTTAGAAAAGATTTC 423  
QY 437 AAGAAAAGATGCGCAACCGACCGCGCGGTGCTGGAATCTGCCATCGCGTCATGTTG 496  
Db 424 AAGAAAATATGCGAACAACCAAGCCAGAGCCTTTGTAATCCGATTCGCGTCATGTAG 483  
QY 497 ATCATCTCGATCGCTGAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 556  
Db 484 AAATTCATGATCGCTTAACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 543  
QY 557 TCCTCGCTGTTAGTCTTATTCGTTTCGTCGCAACAGATCGATTCGATTCGATTCGATTCG 616  
Db 544 TCCTAGCGGTTGATCTTATTCGTTTACTGCTTAAGCAATTAACAGCCTTTACATTTAG 603  
QY 617 GGATCACGAAGCGGTGTTGCGCGACGGGCGAGTAAATCCGCCATTTGTTTAGTTC 676  
Db 604 GCATTACAGAAGCAGGTGGCGCACGCGTGTGCAGTAAATCTGCAGTGGGTTTAGGAA 663  
QY 677 TCGTCTGCTGAAGCATCGCGCACACGCTGCGCGTATCGCTGGCGGCGGATCCGCTCG 736  
Db 664 TGTATTAGCTGAGGCGATTTGGCGATACATACGCGTCTCTTTGGCGGCGAGATCCCTG 723  
QY 737 AAGAGATCAAGTCGCTTTCGATTTTGAATTCGCTGCGTATCCGTTCCGCGAGGATCA 796  
Db 724 AGAAATCAAGTCGCTTTCGATTTTGAATTCGCTGCGTATCCGTTCCGCGAGGATCA 783  
QY 797 ACTTCATCGCTGCGCGACCTGTTTCGCTGAGGAAATTTGATTTATCGGTACGTTAAACG 856  
Db 784 ACTTTATGCTGCCCAACCTGTTCTGCCAAGAAATTTGATTTATCGGTACAGTAAATG 843  
QY 857 CGCTGAGCAACCGCTGGAAGATATCATCTCCGATGAGACGTTTCGATTCGCTGCGTGC 916  
Db 844 CGTAGAACAACCGCTTGAAGATTTATTTACACCAATGGATGTTATTTATTCGTTGTG 903  
QY 917 TGGTGAATGCGCCAGGTGAGGCGCTGTTTCTACCTCGCGCTCACCGCGCAACAAGA 976  
Db 904 TAGTGAATGCTCGCGAGGCGCTGTTTCTCCGATTCGCGCTGCGCGGCTTAACAAA 963  
QY 977 AAAGCGGCTCTATGAAGATGGCGTGCGC---AAAGACCGCTCTGGACAAACAGATATGA 1033  
Db 964 AAAGCGGTTATTATCTTTGCGGAGACGCCAAAGAGCGTTTGTATACGAGATATAG 1023  
QY 1034 TCACACGCTGGAAGCAGCAGTTCGTGCGAAGCC 1069  
Db 1024 TGAACCAATTTAGAAGCAAAATTCGTGCGAAGTC 1058

RESULT 9

US-08-827-190-10/c  
; Sequence 10, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/827,190  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G. 32,837  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 886 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; US-08-827-190-10  
  
Query Match 47.8%; Score 534.6; DB 2; Length 886;  
Best Local Similarity 75.3%; Pred. No. 8.2e-159;  
Matches 666; Conservative 0; Mismatches 219; Indels 0; Gaps 0;  
  
QY 83 TCGCGTACAGTCCATGACCAATAGCGTACGACAGAGTGAAGCAACGGTCAATCAAA 142  
Db 886 TTGCTGTCCAATCTATGACGAATACCGCACGACGATGTTGAAGCCACTGTGCGCAAA 827  
QY 143 TCAAGCGCTGAAACGCGTGGCGTGATATCGTCCGTGATCCGTACCGACGATGGACG 202  
Db 826 TCCAATACCTTGAGCGTGTAGTGTGATATCGTCCGCGTGTCTGTTCTACGATGATG 767  
QY 203 CGCGAAGACGCTTCAAACTCATCAACACAGCAGGTTAAACGTGCGCTGGTGGCTGACATCC 262  
Db 766 CAGCAGAGCGCTTTAAATTAATTAAGCAGCGCTGAATGTGCCATTGGTTGCGGATATTC 707  
QY 263 ACTTCGATATCGCATTTGCGCTGAAAGTAGCGGAATACGCGTCGATTTCTGCGTATTA 322  
Db 706 ACTTTGACTACCGTATCGCATGAAAGTGGCTGAATATGCTGTTGACTGCTACCAATTA 647  
QY 323 ACCCTGGCAATATCGTATGAAGACGCTTATTCGATGCTGCTGCTGCGCGGATA 382  
Db 646 ACCAGGTAAATATCGCGAGTGAAGACGCTTATTCGCAAGTCGTTGATGCTGCTCATC 587  
QY 383 AAAACATTCGATCCGATTTGCGCTTAAACCGCGATCGCTGGAAGAAAGATTCGAAGAAA 442  
Db 586 ACAAATTCCTATCCGTATAGGGTCAATGGCGGCTCACTGGAAGAAAGATATCCAGAAA 527  
QY 443 AGTATGGCAACCGACGCGCGAGCGGCTGCTGGAATCTGCCATGCGTCAATGTTGATCATC 502  
Db 526 AATACGCTGAGCAACACCTGAAGCATTTGGTTGAATCAGCAATGCGACATGTTGATATCT 467  
QY 503 TCGATCGCTGAACCTTCGATTCAGTTCAAAGTCAGGTGAAGCGTCAAGCTCTTCCTCG 562  
Db 466 TGGACAGGCTGAATTTTCGATTCAGTTCAAGGTGTTTAAAGCGTCGAGTGTCTTCTTG 407  
QY 563 CTGTTGAGTCTTATCGTTTGTGCGCAAAACAGATCGATCAGCGCTTGCATCTGGGATCA 622  
Db 406 CGGTGCGCTTATCGTTTATTTGGCGCAAAAATTTGATCAACCACTTCACTCGGTATTA 347  
QY 623 CGAAGCGCGTGTGCGCGACGCGGCGAGTAAATTCGCCATTTGGTTAGGTCTGCTGC 682  
Db 346 CAGAAGCGGTTGGGCTCGTCTGTTTCAAGTGAATCAGCAATTTGGTCTTGGTATTTGT 287  
QY 683 TGTCTGAAGGCATCGCGACACGCTGCGGTATCCGTGCGCGCGATCCGTGCGGAGAGA 742  
Db 286 TGGCTGAAGGTATCGCGATACGTTACGTTATCTCACTCGCGGAGATCTCTGTTGAGAAG 227  
QY 743 TCAAGTCGTTTCGATATTTTGAATCCGTCGATCCGTTCGCGAGGATCAACTTCA 802  
Db 226 TGAAGTCGTTTGTGATTTCTAAATCGTTACGGATCCGCTCACGTGSCATCACTTTA 167  
QY 803 TCGCTGCGCGACCTGTTTCGCGTCAGGAATTTGATGTTATTCGGTACGGTTAACCGCTGG 862

Best Local Similarity 74.9%; Pred. No. 1.1e-185;  
Matches 790; Conservative 0; Mismatches 262; Indels 3; Gaps 1;

17 CAATTCAACGTGAGAAATCAACAGTATTTACGTTGGGAATGCGCGATTTGGCGATGTTG 76  
18 CTTATTAGCGTCTGTAATCGACAAAATTTATGTGGGAATGTAACAAATTTGTGGGATG 63  
19 CTCCTATTCGCGTGAATCAATGACAAATCTCGACCACTGATGTGAAGCGACAGTTG 123  
20 ATCAATCAAGCGCTGGAACGGTTGGCGTGTATATGTCGCGTGTATCCGTACCGACGA 196  
21 CTCAATTAATCAATTAGAAGCGTTGGTGCAGATATTTGTCGTGTATCTGTTCACAA 183  
22 TGACGCGGCGAGAAAGCGTTCAAACTCATCAACAGCAGGTAAACGTCGCGTGTGGCTG 256  
23 TGAATGCTGCGGAAGCATTTAAACAAATTAACAAGTGAATGTTCCGCTCGTAGAG 243  
24 ATATTCATTCGACTATCGTATCGCGTTAAAGTCGAGAAATATGAGTGGATGTTTAC 303  
25 GTATTAAACCTGCAATATCGGTATGAAGAGCGTATTCGCAATGTTGTTGACTGTGGCG 376  
26 GTATCAATCTGCGCAACATTTGCTGTAAGATCGGTCGCTGCGTGTGTTGTTGTTGCGC 363  
27 GCGATAAAACATTTCCGATCCGTTATTTGGGTTTAAACCGCGATCGCTGGAAGATCTGC 436  
28 GAGACAAAATATTCGATTCGTTATTTGGTGAATGCGAGGCTTTTGAAGAAAGATTTGC 423  
29 AAGAAAGTATGCGAACCGACCGCGCGCTGCTGGAATCTGCCATGCGTCAATGTTG 496  
30 AAGAAATATGCGAACCAACCGCAGAGCCCTGTTAGATCCGATTCGCTGATGTAG 483  
31 ATCATCTCGATCCGCTGAATCTCGATCAGTCAAAAGTCAAGCGTGAAGCGTCTGACGTC 556  
32 AAATTTCTAGATCTCTTAATCTCGATCAGTCAAAAGTCAAGCGTGAAGCGTCCGATGTAT 543  
33 TCCTCGCTGTGAGTCTTATCTGTTGCTGGCAAAACAGATCGATCAGCGGTTGATCTGG 616  
34 TCTTACGGTTGAATCTTATCGTTTACTGGCTTAAGCAATTAACAGCGCTTTACATTTAG 603  
35 GATACCGAAGCGGTGTTGGCGGCGAGCGGCGAGTAAATTCGCCATTCGTTTGGTTC 676  
36 GCATTACAGAGCAGTGGCGGCGGCTGTTGCGAGTAAATCTGCAGTGGGTTAGGAA 663  
37 TGTCTGTCTGAGAGCATCGGCGACAGCTCGCGGTATTCGTCGCGCGCGATCCGCTG 736  
38 TGTATTAGTGAAGGCAATTTGGCGATACACTACGCGTCTCTTTGGCGGCGATCCTGTAG 723  
39 AAGAGATCAAGTTCGGTTTGGATATTTTGAATTCGTCGCTGATCGTTTCGCGAGGATCA 796  
40 AGAAATCAAGTTCGGTTTGGATATTTTGAATCTTTACGGAATTCGTTCAAGAGGAATTA 783  
41 ACTTCACTCGCTGCCGACCTGTTCCGCTCAGGAATTTTGAATTCGTTACGTCGCTTAACG 856  
42 ACTTATTCGTTGCCCAACCTGTTCTCGCAGAAATTTGATGTAATCGGTACGTAATG 843  
43 CGCTGAGCAACCGCTGGAAGATATCATCTCCGATGGAAGTTCGATATTCGCTGCG 916  
44 CGTAGAACAACCGCTTGAAGATATTAATACCAATGATGATGATATCTATTCGTTG 903  
45 TGTGTAATCGCCAGTGAAGCGCTGTTTCTACACTCGGCTCACCGCGGCGACAGA 976  
46 TAGTGAATGCTCTGCGAGGACCTCGTCTCGATCTCGGCGTAAACGCGGCTTAACAAA 963  
47 AAGCGCGCTCTATGAAGATGCGGTGCGG - - - AAGAGCGCTCTGAGCAACAGCATAGA 1033  
48 AAGCGGTTTATCTTGAAGCGAGAACGCAAAAGAGCGTTTGAATACGAGATATAG 1023  
49 TCAGCAGCTGGAAGCAGCCATTCGTGCGAAGCC 1068



## RESULT 5

US-09-557-884-1

; Sequence 1, Application US/09557884

; Patent No. 6506581

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: The Nucleotide sequence of

; the Haemophilus influenzae Rd Genome, Fragments

; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/2 inch diskette

; COMPUTER: Dell Pentium

; OPERATING SYSTEM: MS DOS v6.22

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/557,884

; FILING DATE: 25-Apr-2000

; CLASSIFICATION: &lt;Unknown&gt;

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/476,102

; FILING DATE: JUN-5-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971

; REFERENCE/DOCKET NUMBER: PB186P3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-309-8504

; TELEFAX: 301-309-8439

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1830121 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 55.6%; Score 622; DB 4; Length 1830121;

Best Local Similarity 74.8%; Pred. No. 1.4e-184;

Matches 794; Conservative 0; Mismatches 265; Indels 3; Gaps 1;

```
QY 10 CAGGCTCCCAATTCACAGTAGAATAATCAACACGATTTTACGTTGGGAATGTCCGATTGGC 69
Db 390972 CAGCCAACTATTNAGCGTGTGNAATCGACAAAATTTATGTGGAAATGTACCAATTGTT 391031

QY 70 GATGTGCTCCCATCGCGGTACAGTCCATGACCAATACGCTACGACAGACAGCTGCAAGCA 129
Db 391032 GGGGATGCGCTATTTCGCGTGAATCAATGACAAATACTCGCACCACTGATGTGGAAGCG 391091

QY 130 ACGGTCAATCAATCAAGCGCTGGAACGCGTTGGCGCTGATATCGCTCGTGTATCCGTA 189
Db 391092 ACAGTTGCTCAATTAATCAATTAGAACGTTGGTGCAGATATTGTCGTGTATCTGTT 391151

QY 190 CCGACGATGAGCGCGCAGAGCGTTCAAACTCATCAACACAGGTTAACTGCGCGTG 249
Db 391152 CCAACAAATGATGCTCGGAGCAATTTAAACAAATTAACAAAGTGAATGTTCGCTC 391211

QY 250 GTGGCTGACATCCACTTCGACTATCGCTGAAAGTAGCGGAATACGCGCTCGAT 309
Db 391212 GTAGCAGATATTCAATTCGACTATCGTATCGCGTTAAAGTGCAGAAATGAGTGGAT 391271

QY 310 TGTCTCGGTATTAAACCTCGGCAATATCGGTATGAAGAGCGGATTTCCGATGTTGTTGAC 369
Db 391272 TGTTCAGGTATCAATCCTGSCAACATTTGGTTCGTGGAAGATCGGTCGCTCGTGTGAT 391331
```

```
QY 370 TGTGCGCGGATATAAAACATTCGGATCCGATATTGGCGTTAAAGCCGGATCGCTGGAAAAA 429
Db 391332 TGTGCGCGGAGACAAAAATATTCCGATTGCTATTGGTGTAAATGACGGCTCTTTTAAAAA 391391

QY 430 GATCTGCAAGAAAAAGTATGCGGAACCGAGCGCGAGCGCTTCTGGAATCTGCGCATCGCT 489
Db 391392 GATTTTCAAGAAAAATATGCGGAACCAACCGCAGAGCCCTTGTAGAAATCCGCAATTGCGT 391451

QY 490 CATGTTGATCATCTCGATCGCTGAACTTCGATTCAGTTCAAAAGTACAGCGTGAAGAGCTCT 549
Db 391452 CATGTAGAAATTTCTAGATCGCTTAACTTCGATCAGTTTAAAGTGAAGCTTAAAGCTTCC 391511

QY 550 GACGCTTCTCCTCGCTGTTGAGTCTTATCGTTTGGCGAAAAACAGATCGATCAGCGCTTG 609
Db 391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAACAGACCTTTA 391571

QY 610 CATCTGGGATCACCGAAGCGGTGTCGCGAGCGAGCGGGGAGTAAATCCCGCATTTGTT 669
Db 391572 CATTTAGGCATTTACAGAAGCAGGTGCGCACGCGTGTGTCAGTAAATCTGCAGTGGT 391631

QY 670 TTAGTCTGCTGCTGCTGTAAGCGCATCGCGACACGCTCGCGGTATCGCTGGCGCGGAT 729
Db 391632 TTAGGAATGTTATTAGCTGAGGCGCATTTGCGGATACACTACGCGTCTCTTTGGCGGAGAT 391691

QY 730 CCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATTCGTCGCTATCGTTCCGCA 789
Db 391692 CTTGTAGAGGAATCAAGTCGCTTTTGAATTTTGAATCTTTACGGATTTGTTCAAGA 391751

QY 790 GGGATCAACTTCATCGCTGCCGACCTGTCGCGTCAGGAATTTGATGTTATCGGTACG 849
Db 391752 GGAATTAACCTTTATGCTTGCCTGCCCAACCTGTTCTCGCAAGAAATTTGATGTAATCGGTACA 391811

QY 850 GTTAAGCGCGCTGAGCAAGCGCTGGAAGATATCATCACTCCGATCGACGTTTCGATTATC 909
Db 391812 GTAAATGCGCTAGAACAAAGCGCTTGAAGATATTTATACCAATGGATGATCTATTATTC 391871

QY 910 GGCTCGTGTGTAATGGCCAGGTGAGGCGCTGGTTCCTACACTCGGCGTCACCGCGCGC 969
Db 391872 GGTTCGTAGTAATGCTGCTGCGGAGGCACTCGCTCCGATCTCGGCGTACGGGCGGT 391931

QY 970 AACAGAAAAAGCGCTCTATGAAGATGCGGTGCGC---AAAGACCGCTCTGGACAAAC 1026
Db 391932 AACAAAAAAAGCGGTTATTATCTTCGCGAGAAACGCCAAAAAAGAGCGTTTGTATACGAA 391991

QY 1027 GATATGATCGACCACTGGAAGCAGCATTCGTGGGAAGCC 1068
Db 391992 GATATAGTGAACCAATTAGAACAAAAATTCGTGCGAAAGTC 392033
```

## RESULT 6

US-09-643-990A-1

; Sequence 1, Application US/09643990A

; Patent No. 6528289

; GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann

; Owen White

; Hamilton O. Smith

; J. Craig Venter

; TITLE OF INVENTION: The Nucleotide sequence of

; the Haemophilus influenzae Rd Genome, Fragments

; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville,

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/2 inch diskette

; COMPUTER: Dell Pentium

Db 1075 TCTATTATTGTTGTGTAGTGAATGCCCGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134  
Qy 961 ACCGGCGGCAACAGAAAACGGCCCTCTATGAAGATGGGTGGCC---AAAGACCGTCTG 1017  
Db 1135 GCTGGCGCGAAAACCAAAAGTGGTTTCTATGAAGATGGCGTTTCGCAAAAAGAGCGTTT 1194  
Qy 1018 GACAAACAGATATGATCGACAGCTGGAAGCAGCAGTTCGTGCGAAAGCCAGTCAGCTG 1077  
Db 1195 GATAATGACAATATTATGATCAGCTTGAGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254  
Qy 1078 GACGAA 1083  
Db 1255 GATGAA 1260

## RESULT 4

US-09-170-187-4  
; Sequence 4, Application US/09170187  
; Patent No. 6383745  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/170,187  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/827,190  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1345 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 175..1272  
; OTHER INFORMATION: /gene="aarC"  
; OTHER INFORMATION: /note="Similar to E. coli GcpE protein listed by GenBank."

US-09-170-187-4

Query Match 56.4%; Score 631.6; DB 4; Length 1345;  
Best Local Similarity 74.5%; Pred. No. 2.4e-189;  
Matches 809; Conservative 0; Mismatches 274; Indels 3; Gaps 1;  
Qy 1 ATGCATAACGAGCTCCAAATTCACGTAAGAAATCAACAGTATTTACGTTGGGAATGTG 60  
Db 175 ATGCATAATGAATCACCGATAAAAGACGTAATCCACCGCAAAATTTATGTAGGTAACGTG 234

Qy 61 CCGATTGGCGATGTGCTCCCATCCCGTACAGTCCATGACCAATACCGGTACGACAGAC 120  
Db 235 CCTATTGGCGATGTGCTCCCATCCCGTACAGTCCATGACCAATACCGGTACGACAGAC 294  
Qy 121 GTCGAAGCAACGGTCAATCAATCAAGGCGGTGGAACGGTTGGCGCTGATATCTCGCT 180  
Db 295 GTTGAAGCACTGTGCGGCAATCAATCACTTGAGCGGTGTAGGTGTTGATATCTCGTCC 354  
Qy 181 GTATCCGTACCGACCATGAGCGCGCAGAAAGGTTCAAACTCATCAAAACAGCAGGTTAA 240  
Db 355 GTGTCTGTTCTTACCATGATGACGAGAAAGCTTTAAATTAATTAAGCAGCGGTGAAT 414  
Qy 241 GTCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTACGGAATAC 300  
Db 415 GTGCCATTGGTTGCGGATATTTCATTTGACTACCGTATCGCGATGAAAGTGGCTGAAT 474  
Qy 301 GCGCTCGATTGTCTGCGTATTAAACCTGGCAATATCGGTAAATGAAGAGGTTATTCGATG 360  
Db 475 GGTGTTGACTGCTACGAAATTAACCCAGGTAATATCGGCGAGTGAAGAGCGTATTTCG 534  
Qy 361 GTGGTTGACTGCTGCGCGGATAAACATTCGATCCGTTATTTGGGTTAAACCCGCGATCG 420  
Db 535 GTCGTTGATGAGTGTGCTGCTATCACAACATTCCTATCCGTTATAGGGTCAATGCGG 594  
Qy 421 CTGGAAAAGATCTGCAAGAAAAGTATGCGCAACCGACCGCCGAGCGTTGCTGGAATCT 480  
Db 595 CTGGAAAAGATATCCAGAAAATACCGTGAGCCAAACACCTGAAGCATTGGTTGAATCA 654  
Qy 481 GCATGCGTCTGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTACGCGTG 540  
Db 655 GCAATGCGACATGTTGATATCTTGGACAGCGCTGAATTTGATCAGTTCAAGGTCACTG 714  
Qy 541 AAAGCGCTGACGCTCTCTCGCTGTGAGTCTTATCGTTCTGCTGCAAAAACAGATCGAT 600  
Db 715 AAAGCGTCGGATGTCTTCTCGCGTGGCTCTTATCGTTTATTTGGCGCAAAAATTTGAT 774  
Qy 601 CAGCGCTTGCATCTGGGGATCACCGAAGCGGTGGTGGCGCAGCGGGGACAGTAAATCC 660  
Db 775 CAACCACTTCACTCGGTATTACAGAAGCGGTGGGCTCGTTCTGGTTCACTGAATCA 834  
Qy 661 GCATGTTTATAGTCTGTCTGTCTGAGGATCGGCGACACGCTGCGGTATCGCTG 720  
Db 835 GCAATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 894  
Qy 721 GCGGCGATCCGTCGAAGAGATCAAGTCGGTTTTCGATATTTTGAATCGCTGCGTATC 780  
Db 895 GCGGCGATCCGTTGAGGAAGTGAAGTGGTTTTCGATATTTTGAATCGCTGCGTATC 954  
Qy 781 CGTTCCGAGGATCAACTTTCATCGCTCCCGACCTGTTTCGCGTCAGGAATTTGATGTT 840  
Db 955 CGCTCAGTGGCATCAACTTTCATCGCTCCCGACCTGTTTCAGCCCAAGAAATTTGATG 1014  
Qy 841 ATCGTACGTTAAACGCGCTGAGCAACCGCTTGAAGATATCATCATCTCCGATGGAAGT 900  
Db 1015 ATTGGTACGTTAAATGCTTTTGGAGCAGCGCTCGAAGATATTATCAGCGCGATGATGC 1074  
Qy 901 TCGATTATCGCTGCGTGGTGAATGCGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960  
Db 1075 TCTATTATTGGTTGTGTAGTGAATGGCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134  
Qy 961 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCG---AAAGACCGTCTG 1017  
Db 1135 GCTGGCGCAAAACCAAAAGTGGTTTCTATGAAGATGGCGTTTCGCAAAAAGAGCGTTT 1194  
Qy 1018 GACAAACAGATATGATCGACAGCTGGAAGCAGCAATTCGTGCGAAAGCCAGTCAAGT 1077  
Db 1195 GATAATGACAATATTATGATCAGCTTGAGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254  
Qy 1078 GACGAA 1083  
Db 1255 GATGAA 1260

Db 481 GCCATGGCTGATGTTGATCATCTCGATCGCTGAACCTTCGATCAGTTCAAAGTCAGCGTG 540  
Qy 541 AAAGCGTCTACGCTCTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
Db 541 AAAGCGTCTACGCTCTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600  
Qy 601 CAGCGCTGATCTGGGGATCACCGAAGCCGGTGGTGGCGCAGCGGGGAGTAAATCC 660  
Db 601 CAGCGCTGATCTGGGGATCACCGAAGCCGGTGGTGGCGCAGCGGGGAGTAAATCC 660  
Qy 661 GCCATTGGTTAGGTCGTCTGCTGCTGAAGGATCGCGACACGCTGCCGATCGCTG 720  
Db 661 GCCATTGGTTAGGTCGTCTGCTGCTGAAGGATCGCGACACGCTGCCGATCGCTG 720  
Qy 721 GCGGCCGATCCGGTCCGAGAGATCAAAAGTCGGTTTCGATATTTGAAATCGTCGGTATC 780  
Db 721 GCGGCCGATCCGGTCCGAGAGATCAAAAGTCGGTTTCGATATTTGAAATCGTCGGTATC 780  
Qy 781 CGTTCGCGAGGGATCAACTTCATCGCTGCCGACCTGTTCCGGTCAGGAAATTTGATGTT 840  
Db 781 CGTTCGCGAGGGATCAACTTCATCGCTGCCGACCTGTTCCGGTCAGGAAATTTGATGTT 840  
Qy 841 ATCGGTACGGTTAACCGCTGGAGCAACCGCTGGAAGATATCATCTCCGATGGACGTT 900  
Db 841 ATCGGTACGGTTAACCGCTGGAGCAACCGCTGGAAGATATCATCTCCGATGGACGTT 900  
Qy 901 TCGATTATCGCTGCGTGGTGAATGCGCCGACCTGTTCCGGTCAGGAAATTTGATGTT 960  
Db 901 TCGATTATCGCTGCGTGGTGAATGCGCCGACCTGTTCCGGTCAGGAAATTTGATGTT 960  
Qy 961 ACCGGCGCAACAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010  
Db 961 ACCGGCGCAACAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010

RESULT 3

US-08-827-190-4  
; Sequence 4, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: US/08/827,190  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1345 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 175..1272  
; OTHER INFORMATION: /gene= "aarC"  
; OTHER INFORMATION: /note= "Similar to E. coli GopE protein listed by GenBank."  
; US-08-827-190-4  
Query Match 56.4%; Score 631.6; DB 2; Length 1345;  
Best Local Similarity 74.5%; Pred. No. 2.4e-189;  
Matches 809; Conservative 0; Mismatches 274; Indels 3; Gaps 1;  
Qy 1 ATGCAATACACGAGCTCCAAATTCACCGTAGAAAATCAACAGTATTTACGTTGGGAATGTG 60  
Db 175 ATGCAATACGATCACCGATATAAAGACGTAATCCACCGAATTTATGTAGTAACTG 234  
Qy 61 CGATTGGCGATGGTCTCCCATCGCGGTACAGTCCATACCAATACGCGTACGAGTAC 120  
Db 235 CCTATTGGCGATGGTCTCCCATTTGCTGTCCTCAATCTATGACGAATACGGGACGCGAT 294  
Qy 121 GTCGAGCAACCGTCAATCAAAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCGT 180  
Db 295 GTTGAAGCACCTGTGGCGCAAAATCCAAATCACTGAGCGTGTAGGTGTTGATATCGTCCG 354  
Qy 181 GTATCGGTACCGACGATGACGCGGAGAAAGCGTTCAAACTCATCAAAACAGAGGTAAAC 240  
Db 355 GTGCTGTCTTACGATGATGATGAGCAAGAGCTTTAAATTAATTAAGCAGCGGTGAAT 414  
Qy 241 GTGCGCGTGGTGGGTGACATCCACTTCGACTATCGCATTCGCGTGAAGTACGCGAATAC 300  
Db 415 GTGCCATTGGTTGGCGATATTCATTTGACTACCGTATCCGATGAAGTGGCTGAATAT 474  
Qy 301 GCGTCGATGTTCTGCGTATTAACCTTGGCAATATCGGTAAATGAAGAGGTATTCGCATG 360  
Db 475 GGTGTGACTGCTACGAAATTAACCCAGGTAATATCGGCAGTGAAGAGCGTATTCGCCAA 534  
Qy 361 GTGGTTGACTGTGCGCGCATATAAACAATTCGATCCGTTATTTGGCGTTAAACGCGGATCG 420  
Db 535 GTGCTTGTAGTGTCTGCTGCTCATCAACAATCTCTATCCGTATAGGGGTCAATGGCGGTCA 594  
Qy 421 CTGGAAAAAGATCTCGAAGAAAAGTATGCGCAACCGACCGCGAGCGGTTGTCTGGAATCT 480  
Db 595 CTGGAAAAAGATCTCGAAGAAAATACGGTGAGCAACACCTGAAGCACTTGGTTGAATCA 654  
Qy 481 GCGATGCGTCAATGATCATCTCGATCCCTGGAATTCGATCGATCAAAAGTACGCGTG 540  
Db 655 GCAATGCGACATGTTGATATCTTGGACAGGCTGAATTTCCGATCAGTTCAAGGTCACTGT 714  
Qy 541 AAAGGCTCTGACGCTCTTCCCTGCTGTGAGTCTTATCGTTTCTGCGCAAAACAGATCGAT 600  
Db 715 AAAGGCTCGGATGTTCTTCTGCGGTGCTCTTATCGTTTATTTGGCGCAAAAATTTGAT 774  
Qy 601 CAGCGGTTGATCTGGGGATCACCGAAGCCGGTGGTGGCGCAGCGGGGAGTAAATCC 660  
Db 775 CAACCACTTCACCTCGGTATTACAGAGCGGTGGGGCTCGTTCTGGTTCAAGTAAATCA 834  
Qy 661 GCGATTGGTTAGGTCGTCTGCTGCTGAAGGATCGCGGACACCGCTGCCGATATCGCTG 720  
Db 835 GCAATTTGGTCTTGGTATGTTGTTGGCTGAAGGATACGCGGATACGTTACTCTCACTC 894  
Qy 721 GCGGCGGATCCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTGGGTATC 780  
Db 895 GCGGCGGATCCGTTGAGGAAGTGAAGTCGGTTTGTGATATTTCTAAAATCGTTACCGGATC 954  
Qy 781 CGTTCGCGAGGATCAACTTCATCGCCTGCCGACCTGTTCCGCGTCAGGAATTTGATGTT 840  
Db 955 CGCTCAGCTGGGATCAACTTTATTCTTGGCCCAACCTGTTTACGCGCAAGAAATTTGATG 1014  
Qy 841 ATCGGTAGGTTAACGCGCTGGAGCAACCGCTGGAGATATCATCTCCGATGAGCGTT 900  
Db 1015 ATTGGTACGGTAAATGCTTTGGAGCAGCGCTCGAAGATATTTATCACGCGGATGATGTC 1074  
Qy 901 TCGATTATCGGCTGCGTGGTGAATGCGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTC 960

```
Db 61 CCGATTGGCGATGGTCTCCATCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
Qy 121 GTGGAAGCAAGGTCAATCAATCAAGCGCGTGAACCGCGTGGCGTGTATATCGTCGGT 180
Db 121 GTGGAAGCAAGGTCAATCAATCAAGCGCGTGAACCGCGTGGCGTGTATATCGTCGGT 180
Qy 181 GTATCCGTACCGAGTGGAGCGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTAAAC 240
Db 181 GTATCCGTACCGAGTGGAGCGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTAAAC 240
Qy 241 GTCCCGTGGTGGTGAATCCACTTCGACTATCGCATTCGCGTGAAGTGAAGTGAATAC 300
Db 241 GTCCCGTGGTGGTGAATCCACTTCGACTATCGCATTCGCGTGAAGTGAAGTGAATAC 300
Qy 301 GGGTGTGATGTCTGGGTATTAACCTGGCAATATCGGTAAATGAAGCGGTATTCGATG 360
Db 301 GGGTGTGATGTCTGGGTATTAACCTGGCAATATCGGTAAATGAAGCGGTATTCGATG 360
Qy 361 GTGTTGACTGTGCGCGGAGTAAACATTCGATCCGATTCGATTCGCGTGAACCGGATCG 420
Db 361 GTGTTGACTGTGCGCGGAGTAAACATTCGATTCGATTCGCGTGAACCGGATCG 420
Qy 421 CTGGAAGCAAGTCTGAAGAGTATGGCGAAGCGGCGGAGCGGTTGCTGGAATCT 480
Db 421 CTGGAAGCAAGTCTGAAGAGTATGGCGAAGCGGCGGAGCGGTTGCTGGAATCT 480
Qy 481 GCGATGGTGTGATGATCATCTGATCGCTGAACTTCGATTCGATTCGATTCGATTCG 540
Db 481 GCGATGGTGTGATGATCATCTGATCGCTGAACTTCGATTCGATTCGATTCGATTCG 540
Qy 541 AAAGCGTCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTGCTGCGAAGAGATCGAT 600
Db 541 AAAGCGTCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTGCTGCGAAGAGATCGAT 600
Qy 601 CAGCGGTGTGATCTGGGGATCACGGAAGCGGTTGTCGCGGAGCGGCGGAGTAAATCC 660
Db 601 CAGCGGTGTGATCTGGGGATCACGGAAGCGGTTGTCGCGGAGCGGCGGAGTAAATCC 660
Qy 661 GCGATGGTGTGATCTGGGGATCACGGAAGCGGTTGTCGCGGAGCGGCGGAGTAAATCC 720
Db 661 GCGATGGTGTGATCTGGGGATCACGGAAGCGGTTGTCGCGGAGCGGCGGAGTAAATCC 720
Qy 721 GCGGCGGATCGGTCGAAGAGATCAAGTGGTTCGATATTTGAAATCGCGGATC 780
Db 721 GCGGCGGATCGGTCGAAGAGATCAAGTGGTTCGATATTTGAAATCGCGGATC 780
Qy 781 CGTTCGCGAGGGATCAACTTCATCGCTGCGGAGCGGTTGTCGCGTGAAGATTTGATGT 840
Db 781 CGTTCGCGAGGGATCAACTTCATCGCTGCGGAGCGGTTGTCGCGTGAAGATTTGATGT 840
Qy 841 ATCGGTACGGTTAAGCGGTGAGAGCAACCGCTGGAAGATATCATCTCGGATGAGGTT 900
Db 841 ATCGGTACGGTTAAGCGGTGAGAGCAACCGCTGGAAGATATCATCTCGGATGAGGTT 900
Qy 901 TCGATTTACGGTGTGGTGAATGCGGAGGAGCGGCTGTTTCTACCTCGGCGTC 960
Db 901 TCGATTTACGGTGTGGTGAATGCGGAGGAGCGGCTGTTTCTACCTCGGCGTC 960
Qy 961 ACCGGCGGCAACGAAGAAAGCGGCTCTATGAAGATGGCGTGGCGAAGA 1010
Db 961 ACCGGCGGCAACGAAGAAAGCGGCTCTATGAAGATGGCGTGGCGAAGA 1010
```

## RESULT 2

```
US-09-170-187-8
; Sequence 8, Application US/09170187
; Patent No. 6381745
; GENERAL INFORMATION:
; APPLICANT: Racher, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-09-170-187-8
```

```
Query Match 90.3%; Score 1010; DB 4; Length 1010;
Best Local Similarity 100.0%; Pred. No. 6.4e-309;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCATAACCGAGTCCCAATTCACGTAGAAAATCAACACGATATTACGTTGGGAATG 60
Db 1 ATGCATAACCGAGTCCCAATTCACGTAGAAAATCAACACGATATTACGTTGGGAATG 60
Qy 61 CCGATTGGCGATGGTCTCCATCCGCTGACCAATCAAGCGCTGGAACCGCTGATATCGTC 180
Db 61 CCGATTGGCGATGGTCTCCATCCGCTGACCAATCAAGCGCTGGAACCGCTGATATCGTC 180
Qy 121 GTCGAAGCAACCGTCAATCAATCAAGCGCTGGAACCGCTGATATCGTC 180
Db 121 GTCGAAGCAACCGTCAATCAATCAAGCGCTGGAACCGCTGATATCGTC 180
Qy 181 GTATCCGTACCGAGTGGAGCGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTAAAC 240
Db 181 GTATCCGTACCGAGTGGAGCGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTAAAC 240
Qy 241 GTGCGCGTGGTGGTGAATCCACTTCGACTATCGCTGCGCTGAAAGTGAAGTAAAC 300
Db 241 GTGCGCGTGGTGGTGAATCCACTTCGACTATCGCTGCGCTGAAAGTGAAGTAAAC 300
Qy 301 GGGCGTGGTGTCTGGGTATTAACCTGGCAATATCGGTAAATGAAGAGCGTATTCG 360
Db 301 GGGCGTGGTGTCTGGGTATTAACCTGGCAATATCGGTAAATGAAGAGCGTATTCG 360
Qy 361 GTGTTGACTGTGCGGCGGATAAAACATTCGATCCGTTATTCGTTAAACCGCGATCG 420
Db 361 GTGTTGACTGTGCGGCGGATAAAACATTCGATCCGTTATTCGTTAAACCGCGATCG 420
Qy 421 CTGGAAGAGATCTGCAAGAGAGTATGCGGAACCGCGCGGCGGCTGCTGGAATCT 480
Db 421 CTGGAAGAGATCTGCAAGAGAGTATGCGGAACCGCGCGGCGGCTGCTGGAATCT 480
Qy 481 GCGATGGTGTGATGATCATCTGATCGCTGAACTTCGATTCGATTCGATTCGATTCG 540
Db 481 GCGATGGTGTGATGATCATCTGATCGCTGAACTTCGATTCGATTCGATTCGATTCG 540
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 13:55:31 ; Search time 83 Seconds  
(without alignments)  
5950.695 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataacaggctcccaat.....ttcagcaggttgaaaaataa 1119

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents NA:\*

- 1: /cgn2\_6/prodata/1/ina/5A COMB.seq:\*
- 2: /cgn2\_6/prodata/1/ina/5B COMB.seq:\*
- 3: /cgn2\_6/prodata/1/ina/6A COMB.seq:\*
- 4: /cgn2\_6/prodata/1/ina/6B COMB.seq:\*
- 5: /cgn2\_6/prodata/1/ina/PCUTS COMB.seq:\*
- 6: /cgn2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1010	90.3	1010	2	US-08-827-190-8
2	1010	90.3	1010	4	US-09-170-187-8
3	631.6	56.4	1345	2	US-08-827-190-4
4	631.6	56.4	1345	4	US-09-170-187-4
5	622	55.6	1830121	4	US-09-557-884-1
6	622	55.6	1830121	4	US-09-643-990A-1
7	619.8	55.4	1079	2	US-08-827-190-7
8	619.8	55.4	1079	4	US-09-170-187-7
9	534.6	47.8	886	2	US-08-827-190-10
10	534.6	47.8	886	4	US-09-170-187-10
11	515.4	46.1	1644	4	US-09-252-991A-6420
12	515.4	46.1	2202	4	US-09-252-991A-6131
13	512	45.1	2088	4	US-09-252-991A-6339
14	401.4	35.9	1137	4	US-09-328-352-3780
15	299	26.7	555	4	US-09-252-991A-6251
16	269	24.0	886	2	US-08-827-190-9
17	269	24.0	886	4	US-09-170-187-9
18	243.8	21.8	4403765	3	US-09-103-840A-2
19	243.8	21.8	4411529	3	US-09-103-840A-1
20	228.8	20.4	543	4	US-09-252-991A-6209
21	55.4	5.0	1083	4	US-09-221-017B-309
22	43.8	3.9	1230025	4	US-09-198-452A-1
23	40.6	3.6	1206	4	US-09-153-599A-2
24	40.6	3.6	2299	4	US-09-153-599A-1
25	38.4	3.4	1665	4	US-09-221-017B-1017
26	37.8	3.4	435	4	US-09-252-991A-14212
27	37	3.3	408	4	US-09-252-991A-1295

28	37	3.3	1257	4	US-09-252-991A-1225	Sequence 1225, Ap
29	37	3.3	2280	4	US-09-252-991A-1270	Sequence 1270, Ap
30	37	3.3	2616	4	US-09-252-991A-1336	Sequence 1336, Ap
31	35.4	3.2	583	3	US-08-998-416-1089	Sequence 1089, Ap
32	35.4	3.2	737	3	US-08-998-416-1039	Sequence 1039, Ap
33	35	3.1	1383	3	US-09-400-742-7	Sequence 7, Appli
34	35	3.1	1383	3	US-08-618-651A-7	Sequence 6, Appli
35	35	3.1	1383	4	US-09-215-252-6	Sequence 121, App
36	34.8	3.1	33312	4	US-08-311-731A-121	Sequence 5372, Ap
37	34.6	3.1	1461	4	US-09-252-991A-5372	Sequence 13090, A
38	34.4	3.1	1476	4	US-09-252-991A-13090	Sequence 444, App
39	34.2	3.1	1251	4	US-09-252-991A-444	Sequence 12207, A
40	34.2	3.1	1293	4	US-09-252-991A-12207	Sequence 12135, A
41	34.2	3.1	1338	4	US-09-252-991A-12135	Sequence 417, App
42	34.2	3.1	1362	4	US-09-252-991A-417	Sequence 12388, A
43	34.2	3.1	1859	4	US-09-252-991A-12388	Sequence 766, App
44	34.2	3.1	2334	4	US-09-620-312D-766	Sequence 1, Appli
45	34	3.0	8910	3	US-08-779-764A-1	

ALIGNMENTS

RESULT 1  
US-08-827-190-8  
; Sequence 8, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medlen & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,190  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1010 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: unknown  
; MOLECULE TYPE: DNA (genomic)  
US-08-827-190-8

Query Match 90.3%; Score 1010; DB 2; Length 1010;  
Best Local Similarity 100.0%; Pred. No. 6.4e-309;  
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGCGATACCAAGGCTCCCAATTCACGTAGAGAAATCAACACGTATTACGTTGGGAATGTG	60
DB	1	ATGCATAACCAAGGCTCCCAATTCACGTAGAGAAATCAACACGTATTACGTTGGGAATGTG	60
QY	61	CCGATTGGCGATGGTGCTCCCATCGGTACAGTCCATGACCAATACGCGTACGACAGAC	120



```
FEATURES
  source      Location/Qualifiers
              1..1010
              /organism="Escherichia coli"
              /mol_type="genomic DNA"
              /db_xref="taxon:32644"

BASE COUNT   238 a      253 c      287 g      232 t

ORIGIN
Query Match      90.3%; Score 1010; DB 6; Length 1010;
Best Local Similarity 100.0%; Pred. No. 7.7e-239;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACCAAGGCTCCAAATTCACATGAGAAATCAACAGGTAATTTACGTTGGGAATGTG 60
   |||
Db 1 ATGCATACCAAGGCTCCAAATTCACATGAGAAATCAACAGGTAATTTACGTTGGGAATGTG 60

QY 61 CCGATTGGCGATGGTGTCTCCCATGCGGTACAGTCCATGACCAATACGCGTACGACAGAC 120
   |||
Db 61 CCGATTGGCGATGGTGTCTCCCATGCGGTACAGTCCATGACCAATACGCGTACGACAGAC 120

QY 121 GTGGAACCAACGGTCATCAAAATCAAGGCGTGAACGCGTGGCGCTGAATGTCGT 180
   |||
Db 121 GTGGAACCAACGGTCATCAAAATCAAGGCGTGAACGCGTGGCGCTGAATGTCGT 180

QY 181 GTATCCGTACCGACGATGAGACGCGGACAGAGCGTTCAAACTCATCAACAGAGTTTAC 240
   |||
Db 181 GTATCCGTACCGACGATGAGACGCGGACAGAGCGTTCAAACTCATCAACAGAGTTTAC 240

QY 241 GTGCGCGTGGTGGTGAATTCATCTGACTATGCAATGGCGTGAAGTAAAGCGGAATAC 300
   |||
Db 241 GTGCGCGTGGTGGTGAATTCATCTGACTATGCAATGGCGTGAAGTAAAGCGGAATAC 300

QY 301 GGGCGCATTTGTCTGCGATTTAAACCTGGGCAATTCGGTAATGAAGAGCGTATTCGCATG 360
   |||
Db 301 GGGCGCATTTGTCTGCGATTTAAACCTGGGCAATTCGGTAATGAAGAGCGTATTCGCATG 360

QY 361 GTGTTGACTGTGCGCGCATTAATAAATTCGATCCGATTCGTAATGGCGTTAAAGCGGATCG 420
   |||
Db 361 GTGTTGACTGTGCGCGCATTAATAAATTCGATCCGATTCGTAATGGCGTTAAAGCGGATCG 420

QY 421 CTGGAATAAAGATCTGCAAGAAAGATAGCGACGACGCGGCGGCTTGTCTGAAATCT 480
   |||
Db 421 CTGGAATAAAGATCTGCAAGAAAGATAGCGACGACGCGGCGGCTTGTCTGAAATCT 480

QY 481 GCCATGGGTGATCTTTGATCATCTGATCGCCTGAACCTTCAATCAATCAAGTCAAGCGTG 540
   |||
Db 481 GCCATGGGTGATCTTTGATCATCTGATCGCCTGAACCTTCAATCAATCAAGTCAAGCGTG 540

QY 541 AAAGCGTCTGACGCTTCTCTGCTGTGAGTCTTATGTTGCTGGCAAAAACAGATCGAT 600
   |||
Db 541 AAAGCGTCTGACGCTTCTCTGCTGTGAGTCTTATGTTGCTGGCAAAAACAGATCGAT 600

QY 601 CAGCGTTGCACTGCGGATCACCGAAAGCGGTGTGGCGGACGCGGCGAGTAAATCC 660
   |||
Db 601 CAGCGTTGCACTGCGGATCACCGAAAGCGGTGTGGCGGACGCGGCGAGTAAATCC 660

QY 661 GCCATTGGTTAAGTCTGCTGCTGCTGATGAGGCAATCGGCGACAGCTGGCGGTATCGCTG 720
   |||
Db 661 GCCATTGGTTAAGTCTGCTGCTGCTGATGAGGCAATCGGCGACAGCTGGCGGTATCGCTG 720

QY 721 GCGGCCGATCCGATCGAAGAGATCAAAAGTCGTTTCGATTTTGAATCGCTGATC 780
   |||
Db 721 GCGGCCGATCCGATCGAAGAGATCAAAAGTCGTTTCGATTTTGAATCGCTGATC 780

QY 781 CGTTGCGAGGAGATCAATTCATGCGCTGCGGACCTGTTGCGGTCAAGAAATTTGATGTT 840
   |||
Db 781 CGTTGCGAGGAGATCAATTCATGCGCTGCGGACCTGTTGCGGTCAAGAAATTTGATGTT 840
```

```
QY 841 ATCGGTAACGATTACCGGCTGAGACCAACCGCTGGAAGATATCATCTCGATGACGTT 900
   |||
Db 841 ATCGGTAACGATTACCGGCTGAGACCAACCGCTGGAAGATATCATCTCGATGACGTT 900

QY 901 TCGATTATCGGCTGCGGTGTGAATGGCCAGGAGGAGGCGCTGTTCTACACTCGGCGTC 960
   |||
Db 901 TCGATTATCGGCTGCGGTGTGAATGGCCAGGAGGAGGCGCTGTTCTACACTCGGCGTC 960

QY 961 ACCGGCGGACCAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010
   |||
Db 961 ACCGGCGGACCAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010
```

Search completed: November 23, 2003, 15:07:56  
Job time : 4356 secs

```

Db      601 CAGCCGTCATCTGGGGATCACCAGACCGGTGTCGCGACGCGGAGCATTAATCC 660
Qy      661 GCCATTGGTTAGGTCCTGCTGCTGTGAAGGATCGGCGACACGCTGCGGTATCGCTG 720
Db      661 GCCATTGGTTAGGTCCTGCTGCTGTGAAGGATCGGCGACACGCTGCGGTATCGCTG 720
Qy      721 GCGGCCGATCCGCTCGAGAGATCAAAAGTCGTTTCATATTTTGAATGCTGCGTATC 780
Db      721 GCGGCCGATCCGCTCGAGAGATCAAAAGTCGTTTCATATTTTGAATGCTGCGTATC 780
Qy      781 CGTTCCGAGAGGATCACTTCATATGCTGCGCCGACCTGTTGCGCTCAGGAATTTGAGTT 840
Db      781 CGTTCCGAGAGGATCACTTCATATGCTGCGCCGACCTGTTGCGCTCAGGAATTTGAGTT 840
Qy      841 ATCGGATCGGTTAAAGCGCTGAGACCAAGCCTGGAAGATATCATCATCCGATGAGCTT 900
Db      841 ATCGGATCGGTTAAAGCGCTGAGACCAAGCCTGGAAGATATCATCATCCGATGAGCTT 900
Qy      901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db      901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGCGTGGCGCAAGA 1010
Db      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGCGTGGCGCAAGA 1010

RESULT 14
LOCUS      AR208328      1010 bp      DNA      linear      PAT 20-JUN-2002
DEFINITION Sequence 8 from patent US 6383745.
ACCESSION  AR208328
VERSION     AR208328.1 GI:21509454
KEYWORDS
SOURCE      Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1010)
            Rather, P.N.
TITLE       Methods of screening for anti-microbial utilizing aarc and
            compositions thereof
            Patent: US 6383745-A 8 07-MAY-2002;
            Location/Qualifiers
            source          1..1010
            /organism="unknown"
BASE COUNT  238 a      253 c      287 g      232 t
ORIGIN
Query Match      90.3%; Score 1010; DB 6; Length 1010;
Best Local Similarity 100.0%; Pred. No. 7,7e-239;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      301 GCGCGATTTGTCTGGGATTTAACTCCGCGCAATATCGTATGAAGAGCGTATTCGATG 360
Qy      361 GTGGTTGACCTGTCGCGCGGATTAATAAATTCGATTCGATTTAGCGTTAAACCGCGATCG 420
Db      361 GTGGTTGACCTGTCGCGCGGATTAATAAATTCGATTCGATTTAGCGTTAAACCGCGATCG 420
Qy      421 CTGGAATAAAGATTCGAAGAAAGTATGAGGAAACGACGCGCGAGCTGCTGATCT 480
Db      421 CTGGAATAAAGATTCGAAGAAAGTATGAGGAAACGACGCGCGAGCTGCTGATCT 480
Qy      481 GCCATCGTCATGTTGATCATATCGATCGCTGAACTTCGATCAGTTCAAAAGCAGCGTG 540
Db      481 GCCATCGTCATGTTGATCATATCGATCGCTGAACTTCGATCAGTTCAAAAGCAGCGTG 540
Qy      541 AAGCGCTGACGCTCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Db      541 AAGCGCTGACGCTCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
Qy      601 CAGCCGTCATCTGGGATACACCAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db      601 CAGCCGTCATCTGGGATACACCAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Qy      661 GCCATTGGTTAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Db      661 GCCATTGGTTAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720
Qy      721 GCGGCCGATCCGCTCGAAGATCAAAAGTCGTTTGAATTTTGAATTCGCTGCTATC 780
Db      721 GCGGCCGATCCGCTCGAAGATCAAAAGTCGTTTGAATTTTGAATTCGCTGCTATC 780
Qy      781 CGTTCCGAGAGGATCACTTCATATGCTGCGCCGACCTGTTGCGCTCAGGAATTTGAGTT 840
Db      781 CGTTCCGAGAGGATCACTTCATATGCTGCGCCGACCTGTTGCGCTCAGGAATTTGAGTT 840
Qy      841 ATCGGATCGGTTAAAGCGCTGAGACCAAGCCTGGAAGATATCATCATCCGATGAGCTT 900
Db      841 ATCGGATCGGTTAAAGCGCTGAGACCAAGCCTGGAAGATATCATCATCCGATGAGCTT 900
Qy      901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Db      901 TCGATTATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
Qy      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGCGTGGCGCAAGA 1010
Db      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGCGTGGCGCAAGA 1010

RESULT 15
LOCUS      BD084908      1010 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Methods for screening for antimicrobials utilizing (aarc) and
            compositions thereof.
ACCESSION  BD084908
VERSION     BD084908.1 GI:22630518
KEYWORDS   JP 2001523097-A/6.
SOURCE      unidentified
            unidentified
            unclassified.
REFERENCE   1 (bases 1 to 1010)
            Rather, P.N.
TITLE       Methods for screening for antimicrobials utilizing (aarc) and
            compositions thereof
            Patent: JP 2001523097-A 6 20-NOV-2001;
            CASE WESTERN RESERVE UNIVERSITY
            OS Escherichia coli
            PN JP 2001523097-A/6
            PD 20-NOV-2001
            PF 27-MAR-1998 JP 1998546016
            PR 27-MAR-1997 US 08/827190
            PI PHILIP N RATHER
            PC C1201/68, C07K14/24
            CC Methods for screening for antimicrobials utilizing (aarc) and

```

Matches 1103; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 ATGATATACAGGCTCCATTCACAGTAAACACAGCTATTTCAGTTGGAAATG 60
DB 280346 ATGATATACAGGCTCCATTCACAGTAAACACAGCTATTTCAGTTGGAAATG 280287
QY 61 CCGATTGGCGATGGTGTCCCATGCGCGTACAGTCAATGACCAATACGCTACAGAC 120
DB 280286 CCGATTGGCGATGGTGTCCCATGCGCGTACAGTCAATGACCAATACGCTACAGAC 280227
QY 121 GTGAGACCAACGGTCAATCAATCAAGCGCTGAGACGGCTGGCGCTGATTCCTG 180
DB 280226 GTGAGACCAACGGTCAATCAATCAAGCGCTGAGACGGCTGGCGCTGATTCCTG 280167
QY 181 GTATCCGTACCGACGAGTGAAGCGCGGACGCTTCAATCAATCAAGCGCTTAC 240
DB 280166 GTATCCGTACCGACGAGTGAAGCGCGGACGCTTCAATCAATCAAGCGCTTAC 280107
QY 241 GTGCGCGTGGTGGCTGACATTCACATTCGACTATGCGCTGAGAAAGTAGCGAATAC 300
DB 280106 GTGCGCGTGGTGGCTGACATTCACATTCGACTATGCGCTGAGAAAGTAGCGAATAC 280047
QY 301 GCGCGTATGTCTGCTGATTAACCTTGGCAATATCGGTAATGAAGCGCTATTCGATG 360
DB 280046 GCGCGTATGTCTGCTGATTAACCTTGGCAATATCGGTAATGAAGCGCTATTCGATG 279987
QY 361 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCGCTTAAAGCGCGATTCG 420
DB 279986 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCGCTTAAAGCGCGATTCG 279927
QY 421 CTGGAATAAAGATCTGCAAGAAAGATATGCGACACGCGCGAGCGCTTGTCTGATCT 480
DB 279926 CTGGAATAAAGATCTGCAAGAAAGATATGCGACACGCGCGAGCGCTTGTCTGATCT 279867
QY 481 GCCATGCGTATGTGATTCATCTTCGATCGCTGAACTTGCATGATCAAGTCAAGCTG 540
DB 279866 GCCATGCGTATGTGATTCATCTTCGATCGCTGAACTTGCATGATCAAGTCAAGCTG 279807
QY 541 AAGCGCTACGCTTCTCGCTGCTGATTCGCTTAAAGCTTAAAGCTTAAAGCTTAAAG 600
DB 279806 AAGCGCTACGCTTCTCGCTGCTGATTCGCTTAAAGCTTAAAGCTTAAAGCTTAAAG 279747
QY 601 CAGCGCTTGGATCTGCGGATCACCGAAGCGGCTGCGCGACGCGGCGAGTAAATCC 660
DB 279746 CAGCGCTTGGATCTGCGGATCACCGAAGCGGCTGCGCGACGCGGCGAGTAAATCC 279687
QY 661 GCCATTGGTAAAGTCTGCTGCTGCTGAAAGCATCGGCGACGCTGCGATTCGCTG 720
DB 279686 GCCATTGGTAAAGTCTGCTGCTGCTGAAAGCATCGGCGACGCTGCGATTCGCTG 279627
QY 721 GCGGCGCATCCGCTGCGAAGATCAAGTGGTTCGATATTTTGAATGGCTGCTATC 780
DB 279626 GCGGCGCATCCGCTGCGAAGATCAAGTGGTTCGATATTTTGAATGGCTGCTATC 279567
QY 781 CGTTGCGAAGGATCACTTCATCGCTGCGCGACGCTTGCAGTAATTTGATGTT 840
DB 279566 CGTTGCGAAGGATCACTTCATCGCTGCGCGACGCTTGCAGTAATTTGATGTT 279507
QY 841 ATCGGTAAGGTTAAAGCGCTGAGCAACGCTTGAAGATATCATCTCCGATGACGTT 900
DB 279506 ATCGGTAAGGTTAAAGCGCTGAGCAACGCTTGAAGATATCATCTCCGATGACGTT 279447
QY 901 TCGATTATCGGCTCGCGGTGTAATGGCGCGAGTGAAGCGCTGTTTAACTCCGCGTC 960
DB 279446 TCGATTATCGGCTCGCGGTGTAATGGCGCGAGTGAAGCGCTGTTTAACTCCGCGTC 279387
QY 961 ACCGCGCGCAACAAGAAAGCGCTCTATGAAAGTGGCTGCGCAAGAAGCGCTGAC 1020
DB 279386 ACCGCGCGCAACAAGAAAGCGCTCTATGAAAGTGGCTGCGCAAGAAGCGCTGAC 279327
QY 1021 AACACGATATGATGACACGAGCTGGAAGACGCAATTCGTGCGAAAGCGCACTGAC 1080
DB 279326 AACACGATATGATGACACGAGCTGGAAGACGCAATTCGTGCGAAAGCGCACTGAC 279267

```

QY 1081 GAAGCGCTGCAATTCAGCTTCAGAGTTGAAATAA 1119  
DB 279266 GAAGCGCTGCAATTCAGCTTCAGAGTTGAAATAA 279228

RESULT 13  
AR027993 1010 bp DNA linear PAT 29-SEP-1999  
LOCUS Sequence 8 from patent US 5858367.  
ACCESSION AR027993  
VERSION AR027993.1 GI:593966  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE Unclassified.  
1 (bases 1 to 1010)  
AUTHORS Rather, P.N.  
TITLE Methods for screening for antimicrobials utilizing Aac and compositions thereof  
Patent: US 5858367-A 8 12-JAN-1999;  
FEATURES  
source Location/Qualifiers  
1..1010 /organism="unknown"

BASE COUNT 238 a 253 c 287 g 232 t

Query Match 90.3%; Score 1010; DB 6; Length 1010;  
Best Local Similarity 100.0%; Pred. No. 7.7e-239;  
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGATATACAGGCTCCATTCACAGTAAACACAGCTATTTCAGTTGGAAATG 60
DB 1 ATGATATACAGGCTCCATTCACAGTAAACACAGCTATTTCAGTTGGAAATG 60
QY 61 CCGATTGGCGATGGTGTCCCATGCGCGTACAGTCAATGACCAATACGCTACAGAC 120
DB 61 CCGATTGGCGATGGTGTCCCATGCGCGTACAGTCAATGACCAATACGCTACAGAC 120
QY 121 GTGAGACCAACGGTCAATCAATCAAGCGCTGAGACGGCTGGCGCTGATTCCTG 180
DB 121 GTGAGACCAACGGTCAATCAATCAAGCGCTGAGACGGCTGGCGCTGATTCCTG 180
QY 181 GTATCCGTACCGACGAGTGAAGCGCGGACGCTTCAATCAATCAAGCGCTTAC 240
DB 181 GTATCCGTACCGACGAGTGAAGCGCGGACGCTTCAATCAATCAAGCGCTTAC 240
QY 241 GTGCGCGTGGTGGCTGACATTCACATTCGACTATGCGCTGAGAAAGTAGCGAATAC 300
DB 241 GTGCGCGTGGTGGCTGACATTCACATTCGACTATGCGCTGAGAAAGTAGCGAATAC 300
QY 301 GCGGCGCATTCGCTGCGAAGATCAAGTGGTTCGATATTTTGAATGGCTGCTATC 360
DB 301 GCGGCGCATTCGCTGCGAAGATCAAGTGGTTCGATATTTTGAATGGCTGCTATC 360
QY 361 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCGCTTAAAGCTTAAAGCTTAAAG 420
DB 361 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCGCTTAAAGCTTAAAGCTTAAAG 420
QY 421 CTGGAATAAAGATCTGCAAGAAAGATATGCGACACGCGCGAGCGCTTGTCTGATCT 480
DB 421 CTGGAATAAAGATCTGCAAGAAAGATATGCGACACGCGCGAGCGCTTGTCTGATCT 480
QY 481 GCCATGCGTATGTGATTCATCTTCGATCGCTGAACTTGCATGATCAAGTCAAGCTG 540
DB 481 GCCATGCGTATGTGATTCATCTTCGATCGCTGAACTTGCATGATCAAGTCAAGCTG 540
QY 541 AAGCGCTGAGCTCTTCGCTGCTGATTAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAG 600
DB 541 AAGCGCTGAGCTCTTCGCTGCTGATTAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAG 600
QY 601 CAGCGCTTGCATTCGAGGATCAACGAGCGGCTGAGCGCGACGCGGCGAGTAAATCC 660

```



DGKGVKLDSDNIRLLFSLAGNYLANONPDVHOAKLLEDESKIEFLVLSLDEMTBSA  
KADVLLPETSFMERNIGETWGTASLYLSEKPIEDFERADYDMLRDAKGLVE  
AFBSQDEKOWIEHIEOTRLAMPDNIOTKREHLRKSAPHIAFEANIDP  
ONNPPRPSKIEIFSKRLPDMODPEIPALSHVPAFEGEDUKITAKYPIQLITWCK  
NRANSIOYANPMIOEVOTKLMNPDAKRGISEGSVKINDRGSIIPVEITPRI

Query Match 97.7%; Score 1093.4; DB 1; Length 11521;

Best Local Similarity 98.6%; Pred. No. 2e-259;

Matches 1103; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 ATGCATACCAAGGCTCCAAATTCACGTAAGAAATCAACAGCTATTACGTTGGAGATGTG 60
DB 1204 ATGCATACCAAGGCTCCAAATTCACGTAAGAAATCAACAGCTATTACGTTGGAGATGTG 1145
QY 61 CCGATTGGCGATGTCCTCCATGCGGTACAGTCCATGACCAATAGCGGTACGACGAC 120
DB 1144 CCGATTGGCGATGTCCTCCATGCGGTACAGTCCATGACCAATAGCGGTACGACGACGAT 1085
QY 121 GTCGAAACAACGCTCAATCAAAATCAAGGCGGTGAGCGGTGCGCTGATATGTCCTG 180
DB 1084 GTCGAAACAACGCTCAATCAAAATCAAGGCGGTGAGCGGTGCGCTGATATGTCCTG 1025
QY 181 GTATCCGTCACGAGATGAGCGCGGCAAGAGCTTCAAACTCATCAACAGCAGTTAAC 240
DB 1024 GTCTCCGTCACGAGATGAGCGCGGCAAGAGCTTCAAACTCATCAACAGCAGTTAAC 965
QY 241 GTGCGCGGTGAGGTGATCCATCTTGACATATGCGATTCGCGTGAAGATAGCGGATAC 300
DB 964 GTGCGCGGTGAGGTGATCCATCTTGACATATGCGATTCGCGTGAAGATAGCGGATAC 905
QY 301 GGGGTGATTTGTCGCTATTTAAACCTGCGCAATTCGTAATGAAAGCGTATTCGATG 360
DB 904 GGGGTGATTTGTCGCTATTTAAACCTGCGCAATTCGTAATGAAAGCGTATTCGATG 845
QY 361 GTGCTGACTGTGCGCGCGATTAACCAATTCGATTCGTAATGAAAGCGTATTCGATG 420
DB 844 GTGCTGACTGTGCGCGCGATTAACCAATTCGATTCGTAATGAAAGCGTATTCGATG 785
QY 421 CTGGAATAAATCTGCAAGAAATATAGGGAACCGGACGCGGCGGTGCTGGAATCT 480
DB 784 CTGGAATAAATCTGCAAGAAATATAGGGAACCGGACGCGGCGGTGCTGGAATCT 725
QY 481 GCGATGCGTATGTTGATCATCTGATGCGCTGGAATTCGATCGTTCAAAAGTCAGCGTG 540
DB 724 GCGATGCGTATGTTGATCATCTGATGCGCTGGAATTCGATCGTTCAAAAGTCAGCGTG 665
QY 541 AAAGCGTCTGACGCTTCTGCTGCTGTTGAGTCTTATCGTTTGTGCGCAAAACAGATCAT 600
DB 664 AAAGCGTCTGACGCTTCTGCTGCTGTTGAGTCTTATCGTTTGTGCGCAAAACAGATCAT 605
QY 601 CAGCGCTTGCATCTGGGGATCAACGAAGCGGTGTGCGGCGACGGGGCGATTAATTC 660
DB 604 CAGCGCTTGCATCTGGGGATCAACGAAGCGGTGTGCGGCGACGGGGCGATTAATTC 545
QY 661 GCGATGCTTGAAGTGTGCTGCTGCTGGAAGGATCGGCGACAGCTGCGGTGCTGCTG 720
DB 544 GCGATGCTTGAAGTGTGCTGCTGCTGGAAGGATCGGCGACAGCTGCGGTGCTGCTG 485
QY 721 GCGCGCGATCCGCTGGAAGATCAAAAGTCGTTTGAATATTTTGAATCGTCCGATC 780
DB 484 GCGCGCGATCCGCTGGAAGATCAAAAGTCGTTTGAATATTTTGAATCGTCCGATC 425
QY 781 CGTTCCGAGGAGATCAATCTTATGCGCTGCCGACCTGTTGCGTCAAGAAATTTGATGT 840
DB 424 CGTTCCGAGGAGATCAATCTTATGCGCTGCCGACCTGTTGCGTCAAGAAATTTGATGT 365
QY 841 ATCGGTACGTTTAAACGCGCTGGAACAACGCTGGAATAATCATCATCTCCGATGACGTT 900
DB 364 ATCGGTACGTTTAAACGCGCTGGAACAACGCTGGAATAATCATCATCTCCGATGACGTT 305
QY 901 TCGATTATCGGCTGCTGTTGAATGCGCCAGTGAAGCGCTGCTGTTTCACTCGGCGTC 960
DB 304 TCGATTATCGGCTGCTGTTGAATGCGCCAGTGAAGCGCTGCTGTTTCACTCGGCGTC 245

```

```

QY 961 ACCGGCGGCAACAAGAAAGCGGCTCTGTGAAGATGCGGTGCGCAAGACCGTCTGAC 1020
DB 244 ACCGGCGGCAACAAGAAAGCGGCTCTGTGAAGATGCGGTGCGCAAGACCGTCTGAC 185
QY 1021 AACACGATATGATTCAGCAGCAGTGAAGACCGGATTCGTGCGAAGCAGTCAGCTGAC 1080
DB 184 AACACGATATGATTCAGCAGCAGTGAAGACCGGATTCGTGCGAAGCAGTCAGCTGAC 125
QY 1081 GAAGCGCGTTCGATTCAGCTTCAGCAGCTTTGAAATAA 1119
DB 124 GAAGCGCGTTCGATTCAGCTTCAGCAGCTTTGAAATAA 86

RESULT 12
AP002561/c 296827 bp DNA linear BCT 24-APR-2003
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Kurokawa, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tetsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shingawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
20196780
MEDLINE
PUBMED
10734605
REFERENCE
AUTHORS
Ohnishi, M., Murata, T., Nakayama, K., Kohara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shingawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
Escherichia coli K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
20557356
MEDLINE
PUBMED
11106008
REFERENCE
AUTHORS
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shingawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
20564182
MEDLINE
PUBMED
11111050
REFERENCE
AUTHORS
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kohara, S., Shiba, T., Hattori, M. and
Shingawa, H.
Complete genome sequence of enterohemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
21156231
MEDLINE
PUBMED
11258796
REFERENCE
AUTHORS
5 (bases 1 to 296827)
Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shingawa, H. and
Hayashi, T.
Direct Submision
Submitted (126-JUN-2000) Ken Kurokawa, Osaka University, Genome

```

homologous to E. coli K-12 MG1655"

complement (4405. .4983)

/gene="Z3782"

complement (4405. .4983)

/gene="Z3782"

/function="orf; Unknown function"

/note="No significant matches"

/codon\_start=1

/transl\_table=1

/protein\_id="AAG57629.1"

/translation="MTWIKKGGTIRAIADLSQAKWERRALLINCEKROYGEBSVY  
SPHISEVNVSRRLPHYORADVQNTQIQPGIRLHYITMPQISDRPMMDPKCOLCGA  
WCRCEDQVFLNGLNGQICINDALGNCQNCIACVCFHMATVELTILPAKVINLHANK  
VCKTCXSFLEFQONAOQCICQHRVGRMP"

complement (5259. .6080)

/gene="Z3783"

complement (5259. .6080)

/gene="Z3783"

/function="putative enzyme; Not classified"

/note="Residues 2 to 272 of 273 are 25.26 pct identical to  
residues 6 to 276 of 286 from Genpep 118 :  
gi|5002126|gb|AA037317.1|AF135170.8 (AF135170) dimethyl  
sulfoxide reductase subunit C [Yersinia peccis]"

/codon\_start=1

/transl\_table=1

/product="putative dimethyl sulfoxide reductase subunit C"

/protein\_id="AAG57630.1"

/db\_xref="GI:12516914"

/translation="MHEPLPAPFLLIQASVQWTMLAFIHRHILITSSARPVNRPL  
LLAFLAGLVNLIISTLHGLPRLMAYHLSHSSWSREIIFGLYIALGLSTLWV  
LRKGGQMLLMIAVGVIVDFVCAQVYIHFSITWQHYNTLIMFLGTVLGSLAV  
VFRSGILPQIDALRNCGVATIALVRLVLPMLIGDITANMOJATPLPMLML  
GLPDLITLSQIDRISGMVMPFAVGGCKKNIIPALFGSVMLVGSEVMLRFPVFSIG"

complement (6073. .6510)

/gene="Z3784"

complement (6073. .6510)

/gene="Z3784"

/function="putative enzyme; Not classified"

/note="Residues 4 to 134 of 145 are 61.83 pct identical to  
residues 57 to 186 of 205 from Genpep 118 :  
gi|1787122|gb|AAC73391.1| (AB000191) anaerobic dimethyl  
sulfoxide reductase subunit B [Escherichia coli]"

/codon\_start=1

/transl\_table=1

/product="putative dimethyl sulfoxide reductase subunit B"

/protein\_id="AAG57631.1"

/db\_xref="GI:12516915"

/translation="MNSNHYATLSSCNHCADPCTKNCPTMAMHKRPDGIKRVNT  
DKYCGGCAACSCYGAPOMKQEGKMSKCDPCIDLOAKEOEPVATCPLGAKIFGP  
IDELREKGDVSDYKGLPDSGITPRLNVIRHQAEREAHHN"

complement (6699. .9080)

/gene="Z3785"

complement (6699. .9080)

/gene="Z3785"

/function="putative enzyme; Not classified"

/note="Residues 25 to 792 of 793 are 43.59 pct identical  
to residues 3 to 784 of 785 from Genpep 118 :  
gi|1787121|gb|AAC73380.1| (AB000191) anaerobic dimethyl  
sulfoxide reductase subunit A [Escherichia coli]"

/codon\_start=1

/transl\_table=1

/product="putative dimethyl sulfoxide reductase subunit A"

/protein\_id="AAG57632.1"

/db\_xref="GI:12516916"

/translation="MDPDPDKERTTPQVSRRSFLQATSAITLPPISRAKQSPDA  
SPRYTAAVADVVDYCTKPCGCGCQDIRAMRGVAVQITLIDNLEDPMPIMRACV  
RGSEYAFREVVDPDLCTKPMKVRGSGKGERFISWDAITLIDNLEKRIKTOQGPAS  
VYHGTGAVGSGTSGMDARLNTLTQGVLEHVSIGTAAATPYTVGAASGM  
DTLDTLKVILIMGNPEETIFGHNTYIFQMKQNGRTFYIVADPRISDTVSADQWIP  
LLPTPDNALMDMMVYIISNLNDITDTYTLGFEDNSMEGVANESLVAVFGK  
GSHIKTEBEMKRIITHVAQSIQRLADYATITPAALIQGQPORHICGERTAGSTLL  
SISIKGANGIKGMAVAGGSSNRFVCGPMPENPVQAKSINMMQAAADDAKVTQ

		/product="1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase"	
		/protein_id="AAK53460.1"	
		/db_xref="GI:17978528"	
		/translation="MHNQAPIORKSTRIVGVNPIGDGAPLAVQSMTRTRTDEATVNOIKLERGADIVRVSPTMORRSEFKIKQVAVPVLPADHIFPYRLAKXAEVDCIRINPQNGNERIRIMVDCARDKNIPRIQVNAQSLKEDQEKYGETPALLSARHVDHDLNDFDQKRVASVADVAEVESTRLLKQIDQPLHLGTEGARSQVKSALIGLLSEIGDILRVSLADPVEELKVFQDILKSLRIRSRGINFLACPTCSRQEPDVIQVNALEQRLEDIITPMQVSIIGCVNNGPGEALVSTLQVGTGNKSGLYEDGVRKDRLDNNMIDLEARIKAKASQLEARRIDVQVEK"	
BASE COUNT	272 a	280 c	315 g 252 t
ORIGIN			
Query Match	97.7%	Score 1093.4	DB 1; Length 1119;
Best Local Similarity	98.6%	Pred. No. 1.8e-259;	
Matches 1103;	Conservative 0;	Mismatches 16;	Indels 0; Gaps 0;
QY	1	ATGCATPACGAGGCTCCAAATTCAGTAAAGTAAATCAACAGTATTTACCTTGGGAATGTG	60
DB	1	ATGCATPACGAGGCTCCAAATTCAGTAAAGTAAATCAACAGTATTTACCTTGGGAATGTG	60
QY	61	CCGATTGGCGATGGTCTCCCATGCGCTACAGTCCATGACCAATACGCGTACGACAGAC	120
DB	61	CCGATTGGCGATGGTCTCCCATGCGCTACAGTCCATGACCAATACGCGTACGACAGAT	120
QY	121	GTCGAACCAACGGTCAATCAATCAAGGCGCTGGACGCGTTGGCGCTGATATGTCCTG	180
DB	121	GTCGAACCAACGGTCAATCAATCAAGGCGCTGGACGCGTTGGCGCTGATATGTCCTG	180
QY	181	GTCATCCGTACCGACGATGACGCGGCGAAGCGTTCAACTCATCAACACGAGTTAAC	240
DB	181	GTCATCCGTACCGACGATGACGCGGCGAAGCGTTCAACTCATCAACACGAGTTAAC	240
QY	241	GTCGCGCTGGGCGTGCATCATCTTGCATCATGATGCGTGGTGAAGTAGCGGAATAC	300
DB	241	GTCGCGCTGGGCGTGCATCATCTTGCATCATGATGCGTGGTGAAGTAGCGGAATAC	300
QY	301	GCGGTCGATTGTCTGCGTATTTAAACCTGGCAATATCGGTATGAAAGCGTATTCGATG	360
DB	301	GCGGTCGATTGTCTGCGTATTTAAACCTGGCAATATCGGTATGAAAGCGTATTCGATG	360
QY	361	GTCGTTACCTGTGGCGCGATTAACCAATTCCTGATTCGCTTAAACCGCGGATCG	420
DB	361	GTCGTTACCTGTGGCGCGATTAACCAATTCCTGATTCGCTTAAACCGCGGATCG	420
QY	421	CTGGAATAAAGATTCGCAAGAAATATAGCGAACCGACGCGGCGGCTTGTGATTC	480
DB	421	CTGGAATAAAGATTCGCAAGAAATATAGCGAACCGACGCGGCGGCTTGTGATTC	480
QY	481	GCCATGCGTCAATGTTGATTCATCTGATTCGCTGAACTTCGATCAAGTCAAGCGTG	540
DB	481	GCCATGCGTCAATGTTGATTCATCTGATTCGCTGAACTTCGATCAAGTCAAGCGTG	540
QY	541	AAAGCGTCAACGCTTCTCTGCTGATGAGTCTTATGTTTGGCGAATAACAGATGAT	600
DB	541	AAAGCGTCAACGCTTCTCTGCTGATGAGTCTTATGTTTGGCGAATAACAGATGAT	600
QY	601	CAGCGCTGATCTGCGGAGATCAACGAAAGCGGCTGATGCGGCGGCGAGTAAATTC	660
DB	601	CAGCGCTGATCTGCGGAGATCAACGAAAGCGGCTGATGCGGCGGCGAGTAAATTC	660
QY	661	GCCATGCTTAAAGTCTGCTGCTGCTGTAAGAGCATTCGCGACACGCTGCGGATCATG	720
DB	661	GCCATGCTTAAAGTCTGCTGCTGCTGTAAGAGCATTCGCGACACGCTGCGGATCATG	720
QY	721	GCGCGCCATCCGCTGCGAAGATCAAAAGTGGTTTCGATATTTGAATTCGCTGATC	780
DB	721	GCGCGCCATCCGCTGCGAAGATCAAAAGTGGTTTCGATATTTGAATTCGCTGATC	780
QY	781	CGTTGCGAGGAGTCAATTCATCGCTGCGCGACCTGTTGCGGTCAAGAAATTTGATGTT	840
DB	781	CGTTGCGAGGAGTCAATTCATCGCTGCGCGACCTGTTGCGGTCAAGAAATTTGATGTT	840

QY	841	ATCGGTACGGTTATACCGCGCTGGAGCAACCGCTGGAAGATATCATCTCCGATGACGTT	900
DB	841	ATCGGTACGGTTATACCGCGCTGGAGCAACCGCTGGAAGATATCATCTCCGATGACGTT	900
QY	901	TCGATTATGCGCTGCGTGTGAATGGCCAGGTGAGGCGCTGTTCTACCTCGGCGTC	960
DB	901	TCGATTATGCGCTGCGTGTGAATGGCCAGGTGAGGCGCTGTTCTACCTCGGCGTC	960
QY	961	ACCGGCGGCAACAGAAAAGCGGCTCTATGAAAGATGCGTGGCGAAGACCGTCTGAC	1020
DB	961	ACCGGCGGCAACAGAAAAGCGGCTCTCTATGAAAGATGCGTGGCGAAGACCGTCTGAC	1020
QY	1021	AACACGATATGATGACCAACGCTGGAAGACCGCATTCGTCGCAAGCCGATGCTGAC	1080
DB	1021	AACACGATATGATGATGACCAACGCTGGAAGACCGCATTCGTCGCAAGCCGATGCTGAC	1080
QY	1081	GAAAGCGCTGCAATGACCTTCAGCAGGTTGAAAATAA 1119	
DB	1081	GAAAGCGCTGCAATGACCTTCAGCAGGTTGAAAATAA 1119	

RESULT 11	AE005481/c	11521 bp	DNA	linear	BCT 21-MAR-2001
LOCUS	AE005481/c	11521 bp	DNA	linear	BCT 21-MAR-2001
DEFINITION	Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 100 of 290.				
ACCESSION	AE005481				
VERSION	AE005481.1				
KEYWORDS	GI:12516908				
SOURCE	Escherichia coli O157:H7 EDL933				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.				

REFERENCE  
1 (bases 1 to 11521)  
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grodebeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.  
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7

TITLE  
JOURNAL  
NATURE 409 (6819), 529-533 (2001)

MEDLINE  
21074935  
PUBMED  
11206551

REFERENCE  
2 (bases 1 to 11521)

AUTHORS  
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grodebeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.  
Direct Submission  
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

TITLE  
JOURNAL  
Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

FEATURES

source

1..11521  
/organism="Escherichia coli O157:H7 EDL933"  
/mol\_type="genomic DNA"  
/strain="EDL933"  
/serotype="O157:H7"  
/db\_xref="taxon:155864"  
/note="enterohemorrhagic"  
/complement(86..1204)  
/gene="gcpB"  
/note="synonym: 23778"  
complement(86..1204)  
/gene="gcpB"  
/function="orf; Unknown function"  
/note="Residues 1 to 372 of 372 are 100.00 pct identical to residues 1 to 372 of 372 from Escherichia coli K-12 Strain MG1655: B2515"  
/codon\_start=1

CDs



gene complement (7286, .8551)  
/gene="yfcB"  
/locus\_tag="c2876"  
complement (7286, .8551)  
/gene="yfcB"  
/locus\_tag="c2876"  
/EC\_number="2.1.1.72"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2330; Escherichia coli O157:H7 ortholog: z3593"  
/codon\_start=1  
/transl\_table=1  
/product="Hypothetical adenine-specific methylase yfcB"  
/protein\_id="AAN81326.1"  
/db\_xref="GI:26109124"  
/translation="MAGIGFLINOPLVMPHADPTLRSDNRVILNLASPHOLAKOR LVILAKCFLPHLSISNPTPTSTTDREKRIEPPHPRNADLSSSSQTSRRN YCHALKEEYDKIFVDEAVNELDTIQMLKWSVSRFAANIWGHGTDNPDWADVLY LPLVLYPLDIPEDMRTARLTSSEKRIIVEIRVNERIIVAVLTNRAMPFGHEFYD ERLVLPSPYIGELINNKFAGLISKPOHIDMCTGSGCIIACVAPPAPEADVADIS PDALVAEONIEEGLIHNVIPIRSDLFRLPKVQYDLITNPYVDVADMDLPNEY RHPBELGLASTGDKLTRIRILGNADYLDADVILICEVNSVHLMEOYPDVPTWL

Query Match 98.1%; Score 1098.2; DB 1; Length 300099;  
Best Local Similarity 98.8%; Pred. No. 1,4e-260;  
Matches 1106; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 1 ATGCATACAGGCTCCATTCACAGTAAGAAATCAACAGTATTTACGTTGGAAATG 60  
Db 188410 ATGCATACAGGCTCCATTCACAGTAAGAAATCAACAGTATTTACGTTGGAAATG 188351

Qy 61 CCGATTGGCGATGTCCTCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAG 120  
Db 188350 CCGATTGGCGATGTCCTCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAG 188291

Qy 121 GTCGAACCAACGGTCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATTCGTCGT 180  
Db 188290 GTCGAACCAACGGTCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATTCGTCGT 188231

Qy 181 GTATCCGTCACCGACGATGACGGCGGACGCGTCAACTCATCAACGACGTTAAC 240  
Db 188230 GTATCCGTCACCGACGATGACGGCGGACGCGTCAACTCATCAACGACGTTAAC 188171

Qy 241 GTCGCGCTGTCGTCGTCATCCACTTCGATTCGATTCGATTCGCGTGAAGAGCGAATAC 300  
Db 188170 GTCGCGCTGTCGTCGTCATCCACTTCGATTCGATTCGATTCGCGTGAAGAGCGAATAC 188111

Qy 301 GGGCGTCGATTCGTCGCTATTAAACCTGGCAATATCGGTATGAAGAGCGTATTCGATG 360  
Db 188110 GGGCGTCGATTCGTCGCTATTAAACCTGGCAATATCGGTATGAAGAGCGTATTCGATG 188051

Qy 361 GTGTTGATCTGTGCGCGCCATAAAACATTCCGATCCGTATTTGGCGTTAACGCCGATCG 420  
Db 188050 GTGTTGATCTGTGCGCGCCATAAAACATTCCGATCCGTATTTGGCGTTAACGCCGATCG 187991

Qy 421 CTGAAAAAAGATCTGCAAGAAAGTATGCGAACCGACGCGGAGCGTTCGTAATCT 480  
Db 187990 TTGAAAAAAGATCTGCAAGAAAGTATGCGAACCGACGCGGAGCGTTCGTAATCT 187931

Qy 481 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAATTCGATCATGTTCAAAAGTCAGCGTG 540  
Db 187930 GCCATGCGTCATGTTGATCATCTCGATCGCCTGAATTCGATCATGTTCAAAAGTCAGCGTG 187871

Qy 541 AAAGCGTCTGACGCTCTTCTCGCTGTTGAGCTTTATTCGTTGCTGGCAAAACAGATCGAT 600  
Db 187870 AAAGCGTCTGACGCTCTTCTCGCTGTTGAGCTTTATTCGTTGCTGGCAAAACAGATCGAT 187811

Qy 601 CACCGCTGATCTGGGGATCACGGAAGCGGTCGCGCGGACGGGGAGATTAATCC 660  
Db 187810 CACCGCTGATCTGGGGATCACGGAAGCGGTCGCGCGGACGGGGAGATTAATCC 187751

Qy 661 GCCATTGTTAAGTCTGCTGTCGTAAGGATCGGCGACAGCGTCGCGTATCGCTG 720  
Db 187750 GCCATTGTTAAGTCTGCTGTCGTAAGGATCGGCGACAGCGTCGCGTATCGCTG 187691

Db 187750 GCTATTGTTTAAAGTCTGCTGCTGTCGTAAGGATCGGCGACAGCGTCGCGTATCGCTG 187691

Qy 721 GCGCGCATCCGCTCGAAGATCAAAAGTCGTTGATATTTGAATGCTGCGATC 780  
Db 187690 GCGCGCATCCGCTCGAAGATCAAAAGTCGTTGATATTTGAATGCTGCGATC 187631

Qy 781 CGTTTCGAGGATCACTTCATCGCTGCGCCAGCTGTTGCGTCGAGAAATTTGATGTT 840  
Db 187630 CGTTTCGAGGATCACTTCATCGCTGCGCCAGCTGTTGCGTCGAGAAATTTGATGTT 187571

Qy 841 ATCGTATCGGTTAACCGCTGAGCAACGCTGGAAGATATCATCTCCGATGACGTT 900  
Db 187570 ATCGTATCGGTTAACCGCTGAGCAACGCTGGAAGATATCATCTCCGATGACGTT 187511

Qy 901 TCATATATGCGCTGCGTGGTGAATGCGCCAGGAGAGCGCTGGTTCTACCTGCGCGTC 960  
Db 187510 TCATATATGCGCTGCGTGGTGAATGCGCCAGGAGAGCGCTGGTTCTACCTGCGCGTC 187451

Qy 961 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGCGGTGCGCAAGACCGTCTGAC 1020  
Db 187450 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGCGGTGCGCAAGACCGTCTGAC 187391

Qy 1021 AACACGATATGATGACACGACTGGAAGACGCAATTCGTGCGAAGCCAGTCAGTGGAC 1080  
Db 187390 AACACGATATGATGATGACACGACTGGAAGACGCAATTCGTGCGAAGCCAGTCAGTGGAC 187331

Qy 1081 GAAGCGCGTGAATGACGTTGACGCTTGAAGAAATTA 1119  
Db 187330 GAAGCGCGTGAATGACGTTGACGCTTGAAGAAATTA 187292

RESULT 10  
LOCUS AY033515 1119 bp DNA linear BCT 26-DEC-2001  
DEFINITION Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (ispG) gene, complete cds.  
ACCESSION AY033515  
VERSION AY033515.1 GI:17978527  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 1119)  
AUTHORS Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A., Arigoni, D. and Roidich, F.  
TITLE Studies on the nonmevalonate pathway to terpenes: The role of the Gcpe (ispG) protein  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)  
PUBMED 11752431  
REFERENCE 2 (bases 1 to 1119)  
AUTHORS Roidich, F., Hecht, S., Adam, P., Bacher, A. and Eisenreich, W.  
TITLE Direct Substitution  
JOURNAL Submitted (27-Apr-2001) Institut fuer Organische Chemie und Biochemie, Lehrstuhl III, Technische Universitaet Muenchen, Lichtenberg Strasse 4, Garching D-85747, Germany  
FEATURES  
source  
1..1119  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
1..1119  
/gene="ispG"  
1..1119  
/gene="ispG"  
/function="converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate, last but one step of the mevalonate-independent pathway for isoprenoid biosynthesis"  
/note="ispG"  
/codon\_start=1  
/transl\_table=11



gene	complement (163. .1383) /gene="fabB" /locus_tag="c2869" complement (163. .1383) /gene="fabB" /locus_tag="c2869" /EC_number="2.3.1.41" /function="enzyme; Fatty acid and phosphatidic acid biosynthesis" /note="Escherichia coli K-12 ortholog: b2323; Escherichia coli O157:H7 ortholog: z3586" /codon_start=1 /transl_table=1 /product="3-oxoacyl-[acyl-carrier-protein] synthase I" /protein_id="AA081319.1" /db_xref="GI:26109117" /translation="MKSAVITGLGIYSSINNOEVLASLBERSGITTSQELKDSGM RSHVGNVVKLDITGLIDRVYRMSDASIFAFLSMEQIADAGLSPEAYONPRVGLI AGSGGSPRFQVFGADAMRGRGLKAVGPVYVKAMSGVSACTAPFKIHGVYIS SACATSAHICGNAVEIOIGKODIVFAGGEBELCWMACEPDMGALSTKYNDPEKA SRVYDARDGSPVLAGSGMVEBELHIALRGAIYAYVYCGATSGDAMVPSGEG AYRCMKRMAHGVDPIDYLNSHGTSITVGVKELAIRVFGDKSPAIKTKYMTGHS LGAGVGEALYSLMLEHGFVADSIINIELEDEQAAGLIVTETTDRELTLVMSNFGF GGTNATLMRKLD"
gene	1482. .3548 /gene="yfcK" /locus_tag="c2870" 1482. .3548 /gene="yfcK" /locus_tag="c2870" /function="putative" /note="Escherichia coli K-12 ortholog: b2324; Escherichia coli O157:H7 ortholog: z3587" /codon_start=1 /transl_table=1 /product="Hypothetical protein yfcK" /protein_id="AA081320.1" /db_xref="GI:26109118" /translation="MBSLTHIAKIEPPRVRYVYKYSTOPANLENACTPVSRDF DDVFSNDNGLEETRYFLDGNHLEAFPPHPHPLFVASGFGTGNFTLTMQAPDQ FRAHPQAOLOQLRHFISFEKPEPLTRADLALAHQWPELPAWAOLOQWPELPGCH RLLDEGHVTLDLWFDGIDNELTSQDLSDIALAHQWPELPAWAOLOQWPELPGCH RLLARPGSTLATFTSAGVFRGLQAGFTMQRKGFGRKEMLCGVMEQTLPPFCSTP FNMHSGSKOBAITIGGIIASALISLALRGMQVTLVCADEAPALASGNGROAIVPL LSHDEALNRFNSNAPFARFPYDLIPVKEDHMCSTQCGMDEKSHKIAQMLNDL PAELAVANAVAEQITGVANCSGITVPGGMVCEPMLTRNVELAQQGLQSHIYH QLODLSRKDGWMLNFAQDQAHSHVVLANGHQISGFSQTSLLPVYSAVGQVSHIPT TPELAKIKOVLICYGYLTPQNPANQDHICASVYHSGSEETAYDEDOQORBLIDCF PQVQMAKEVYSGKEARGVRCATRDPLPMVGANPYDADTLVEYASLAEKKDAVSP VYDDLPMFALGSRGLCSAPLCAEILAAQSEBPIPMDDASTLAALNNRLVWKKLKG KAYKAG"
gene	complement (3669. .3947) /gene="yfcL" /locus_tag="c2871" complement (3669. .3947) /gene="yfcL" /locus_tag="c2871" /note="Escherichia coli K-12 ortholog: b2325; Escherichia coli O157:H7 ortholog: z3588" /codon_start=1 /transl_table=1 /product="Hypothetical protein yfcL" /protein_id="AA081321.1" /db_xref="GI:26109119" /translation="MIAFESRIALLIDGAVDHAASDELFAAGYLRLGLTLTAIELES GDHSAQAVHTVVSQSLERKAGIELSPROALVTDMWEHLFOASQO"
gene	complement (3981. .4529) /gene="yfcM" /locus_tag="c2872" complement (3981. .4529) /gene="yfcM" /locus_tag="c2872" /function="putative"
CDS	
gene	/note="Escherichia coli K-12 ortholog: b2326; Escherichia coli O157:H7 ortholog: z3589" /codon_start=1 /transl_table=1 /product="Hypothetical protein yfcM" /protein_id="AA081322.1" /db_xref="GI:26109120" /translation="MNSTHVEQLIEIFNSCFADENFTLTKGDEPIYLPADAEPVY NRIVFAHGYSAAIHIEISHWCIGKARBDLVDFGWYCPDGDADOTSGOFEEVEXPO ALDMFCVACAGYAPNVSCDNLEGPDEPDVYQRRVHAQVMYLIANGIDEPARAFIKA LQNYHTPELTLEQFPWPBALN"
gene	complement (4529. .5338) /gene="yfcA" /locus_tag="c2873" complement (4529. .5338) /gene="yfcA" /locus_tag="c2873" /function="putative structure" /note="Escherichia coli K-12 ortholog: b2327; Escherichia coli O157:H7 ortholog: z3590" /codon_start=1 /transl_table=1 /product="Hypothetical protein yfcA" /protein_id="AA081323.1" /db_xref="GI:26109121" /translation="METSFLPMVSPFLIGVLPFVAMLAGFIDSIAGGGGLTLTALM AAGSPANALATNKLQACGGSATITVFRRVVSLSDCKINIMFVMSGALLVO YVQADYLRQILPIVLCIGLFLPLMPKLGEBERQRRMYLPPALLAGGCVGYDFG PAAGSFYALAFVTLGCFNLAKATAHAKLANATSNIGGLFTLDGKRVATGFWLVG QFLGARMSRLVLSKQKLRPMIVYVSAMSAKLLYDSHGGEILHMLGMN"
gene	complement (5338. .6162) /gene="mepA" /locus_tag="c2874" complement (5338. .6162) /gene="mepA" /locus_tag="c2874" /function="enzyme; Murein sacculus, peptidoglycan" /note="Escherichia coli K-12 ortholog: b2328; Escherichia coli O157:H7 ortholog: z3591" /codon_start=1 /transl_table=1 /product="Penicillin-insensitive murein endopeptidase precursor" /protein_id="AA081324.1" /db_xref="GI:26109122" /translation="MNKTIALLALIASASLAATFWQKITPVPGSAOSIGSFSGC IYGAIDLPIQSHYQVMRYDORRYFCHPLVNFIOFLSRQVSNLGMGTVLIGMGMPA GGRFNGHSHQGLDVIDIFLQPKTRWTSQDLRPOALDLYSRDGHVVPALMKPEI FSLIKLAAQDKDVTRLFVNPAIKOOLCLAGTDRLMLRVRRPMFQRAHMYLRCPA DBLECEDQPLPPRGDCCGAELQSWFPPRPKGTTPKPKTPPLPSPCCALLDBHTV"
gene	complement (6166. .7254) /gene="aroC" /locus_tag="c2875" complement (6166. .7254) /gene="aroC" /locus_tag="c2875" /EC_number="4.2.3.5" /function="enzyme; Amino acid biosynthesis: Chorismate" /note="Escherichia coli K-12 ortholog: b2329; Escherichia coli O157:H7 ortholog: z3592" /codon_start=1 /transl_table=1 /product="Chorismate synthase" /protein_id="AA081325.1" /db_xref="GI:26109123" /translation="MMAAGNTIGQLFRVTTFGESHGALCIVDGVPPGPIPLTEADLOH DDBRRPGTSRTTORBPDOVKILISGVEGYTGTSLILENTDORSODYSALDV FRPGHADYVYKQVGLRDYRGGGSRARETARVAAGALAKKYLAAPKPIEIRGLCTO MEDIPLEIKDMSQVEQNPFFCPDPKIDLDLDELMLRKKEGSGIKATKYVVSAGAG LGEVPVRLDADIDIAHLKMSINAVKGEIDGDVVALRQSQRNDEITKGFQSNHAGG IILGSISSGOOITAHMALKPTSSITVPGRTINFGEEVENITGKRHDPVCVIRAVPIAE AMLAIVLMHLLRQARNADVKTIDIPRW"
CDS	

repeat\_region  
10137..10170  
/note="REP (repetitive extragenic palindromic) element, contains 1 REP sequence"  
complement(10179..11462)  
/gene="pepB"  
/locus\_tag="S2742"  
complement(10179..11462)  
/gene="pepB"  
/locus\_tag="S2742"  
/function="putative enzyme, Not classified"  
/note="residues 1 to 427 of 427 are 92.50 pct identical to residues 30 to 456 of 456 from Escherichia coli K-12 : B5233"  
/codon\_start=1  
/transl\_table=1  
/product="putative peptidase"  
/protein\_id="A01789.1"  
/db\_xref="GI:30042169"  
/translation="MTAMKLTITSTPADRMSEKATYSINDDITLHLNGADLGI  
QPARKIDLGKIHVLSGGMADRCMAFMQGYKGRKVEWPLDDAOEOLD  
RLMIDVRDTINAPAEELGPSOLAQAVULISVADRYTATITKEDIPEGYML  
HTYGRGSRSPVLADYNTGDKKAVYACLVKGTTFPSGGSTIKQTFPMQMSD  
MGAATVGLALFMRKLRVLFCLCADNLSGNAFLGDIITRYNRKVEVMT  
DAGRLVLADGLDASAKPEMIIDATLTGAATLGNVHALFSPDDLAERLLAS  
ASQENEPFRLPLAEFRSOLPSNFAELNNTGSAAYPAGASTAAGFLSHFVENYQGM  
LHIDCSATYRKAPVEQWSAGATGLGVRTIANLTL7A"  
complement(11521..11721)  
/gene="yfhJ"  
/locus\_tag="S2743"

Query Match 98.2%; Score 1098.4; DB 1; Length 290380;  
Best Local Similarity 99.0%; Pred. No. 1.3e-260;  
Matches 1105; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 ATGCATACCAAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTCGTTGGGAATGTG 60  
Db 1209 ATGCATACCAAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTCGTTGGGAATGTG 1150

Qy 61 CCGATTGGCGATGGTGTCCCATGCGCGTACACGTCATGACCAATACGCGTACGACGAC 120  
Db 1149 CCGATTGGCGATGGTGTCCCATGCGCGTACACGTCATGACCAATACGCGTACGACGAC 1090

Qy 121 GTCGAACCAACGTCATCAATCAAGCGCGTGAACGCGTGGCGTATGTCGCGT 180  
Db 1089 GTCGAACCAACGTCATCAATCAAGCGCGTGAACGCGTGGCGTATGTCGCGT 1030

Qy 181 GTATCCGTACCGACGATGACGCGCGCAAGACGTTCAACTCATCAACACGAGTTTAC 240  
Db 1029 GTCTCCGTACCGACGATGACGCGCGCAAGACGTTCAACTCATCAACACGCGTTTAC 970

Qy 241 GTGCGCGTGTGGTGTGACATCCATTTGACATATGCGATGGCGTGAAGTACGGAATAC 300  
Db 969 GTGCGCGTGTGGTGTGACATCCATTTGACATATGCGATGGCGTGAAGTACGGAATAC 910

Qy 301 GGGCGTCAATGTCTGCGTATTAAACCTGGCAATATGCGTATGAAGAGCGTATTCGATG 360  
Db 909 GGGCGTCAATGTCTGCGTATTAAACCTGGTATATGCGTATGAAGAGCGTATTCGATG 850

Qy 361 GTGCTTACTGTGGCGCGATTAACCATTCGATTCGATTTGCGTTAAACGCGGATTCG 420  
Db 849 GTGCTTACTGTGGCGCGATTAACCATTCGATTCGATTTGCGTTAAACGCGGATTCG 790

Qy 421 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACGCGCGCGAGCGGTGCTGGAATCT 480  
Db 789 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACGCGCGAGCGGTGCTGGAATCT 720

Qy 481 GCGATGCGTATGTTGATCTCGATGCGCTGAACCTTCGATCAAGTACAGCGTG 540  
Db 729 GCGATGCGTATGTTGATCTCGATGCGCTGAACCTTCGATCAAGTACAGCGTG 670

Qy 541 AAAGCGTCTACGCTTCTCGCTGTGATGCTTATGCTTGTGCGCAAAACAGATCAT 600  
Db 541 AAAGCGTCTACGCTTCTCGCTGTGATGCTTATGCTTGTGCGCAAAACAGATCAT 600

Db 669 AAAGCGTCTACGCTTCTCGCTGTGATGCTTATGCTTGTGCGCAAAACAGATCAT 610  
Qy 601 CAGCCGTCATCTGGGGATCACGGAAGCGGTGTGCGCGACGCGGACGATTAATCC 660  
Db 609 CAGCCGTCATCTGGGGATCACGGAAGCGGTGTGCGCGACGCGGACGATTAATCC 550

Qy 661 GCCATTGTTTAAAGTCTGCTGCTGTGGAAGGATCGCGACACGCTGCGATGCTG 720  
Db 549 GCCATTGTTTAAAGTCTGCTGCTGTGGAAGGATCGCGACACGCTGCGATGCTG 490

Qy 721 GCGCGCGATCCGTCGGAAGATCAAAAGTGGTTTGATATTTTGAATGCTGCGATC 780  
Db 489 GCGCGCGATCCGTCGGAAGATCAAAAGTGGTTTGATATTTTGAATGCTGCGATC 430

Qy 781 CGTTCCGAGAGGATCAACTCATCGGCTGCGCGACCGTTCGCGTACGGAATTTGATGT 840  
Db 429 CGTTCCGAGAGGATCAACTCATCGGCTGCGCGACCGTTCGCGTACGGAATTTGATGT 370

Qy 841 ATCGTACGTTTAAACGCGCTGAGCAACGCTGGAAGATTCATCACTCCGATGACGTT 900  
Db 369 ATCGTACGTTTAAACGCGCTGAGCAACGCTGGAAGATTCATCACTCCGATGACGTT 310

Qy 901 TCGATTATCGGCTGCGTGTGGAATGCGCCAGGTGAGCGCGTGTTCATCACTCGCGTC 960  
Db 309 TCGATTATCGGCTGCGTGTGGAATGCGCCAGGTGAGCGCGTGTTCATCACTCGCGTC 250

Qy 961 ACCGGGGCAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGCGCTCGAC 1020  
Db 249 ACCGGGGCAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGCGCTCGAC 190

Qy 1021 AACACGATATGATTCGACGACGCTGGAAGACGCAATTCGTGCGAAACGATCGATGAC 1080  
Db 189 AACACGATATGATTCGACGACGCTGGAAGACGCAATTCGTGCGAAACGATCGATGAC 130

Qy 1081 GAAGCGCTCGAATTGACGTTGACAGGTTGAAAAA 1116  
Db 129 GAAGCGCTCGAATTGACGTTGACAGGTTGAAAAA 94

RESULT 9  
AE016764/c 300099 bp DNA linear BCT 24-DEC-2002  
LOCUS  
DEFINITION  
Escherichia coli CF703 section 10 of 18 of the complete genome.  
ACCESSION  
AE016764.1 GI:26109116  
VERSION  
AE016764.1 GI:26109116  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli CF703  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1 (bases 1 to 300099)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Domeneberg,M.S. and Blattner,F.R.  
Extensive Mosaic Structure Revealed by the Complete Genome Sequence  
of Uropathogenic Escherichia coli  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
JOURNAL  
PUBMED  
12471157  
REFERENCE  
2 (bases 1 to 300099)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Domeneberg,M.S. and Blattner,F.R.  
Direct Submision  
Submitted (20-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
TITLES  
JOURNAL  
FEATURES  
source  
1..300099 Location/Qualifiers  
/organism="Escherichia coli CF703"  
/mol\_type="genomic DNA"  
/strain="CF703"  
/db\_xref="taxon:199310"

TITLE Direct Submission  
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
Source  
1. .290380  
/organism="Shigella flexneri 2a str. 2457T"  
/mol\_type="genomic DNA"  
/strain="2457T"  
/serotype="2a"  
/db\_xref="taxon:198215"  
complement (82. .1209)  
/gene="gcpB"  
/locus\_tag="S2733"  
complement (82. .1209)  
/gene="gcpE"  
/locus\_tag="S2733"  
/note="residues 1 to 372 of 375 are 96.23 pct identical to  
residues 1 to 372 from Escherichia coli K-12 :  
B2515"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAPI7888.1"  
/db\_xref="GI:30042163"  
/translation="MHNQAPIORRKSRIYVGNVPIGDGAPIAVOSMTNRTTIDVEAT  
VNOIKALERVGADIVRVSVPTMDAAEFKLIKORVNVPLVDIHPDRIALKYAEVGV  
DCIRINGNIGNERIRIMVYDCARDKNIPIRIGVMSGLEKDIQKXGEPTRALLFS  
AMHVDHLDRNFQKRVKASVPLAVESYRLAKQIDQPHLGITTEGARGAV  
KSAIGLGLLESGIDTLRVSLADPVEELKVGEDLIKSLIRSRGINFLACPTCSQ  
EPFVIGTVNLEGLIEDIITPMDSIIGCVNNGEALVSTLGVGTGNNKSGLYEDGV  
RKRLDNDMDIDLEARIKASOLSDEARRIDVOVGKITT"  
complement (1236. .2249)  
/gene="yfgA"  
/locus\_tag="S2734"  
complement (1236. .2249)  
/gene="yfgA"  
/locus\_tag="S2734"  
/function="putative membrane; Not classified"  
/note="residues 1 to 337 of 337 are 80.71 pct identical to  
residues 1 to 337 of 337 from Escherichia coli K-12 :  
B2516"  
/codon\_start=1  
/transl\_table=1  
/product="putative membrane protein"  
/protein\_id="AAPI7889.1"  
/db\_xref="GI:30042164"  
/translation="MNTTEATHDQNEALTTGARLNAREQLGSGQAAVERLCLKVSTV  
RDIEEDKAPADLSTFLRGYIRSYARLVHIDPEELPGLKQAPLRAKYAPQSFSL  
GKRKKRDGWLMTFTLVLFVIGLSGAMMMQDHKAQOERTTMADQSSAELSSNSFO  
GSVPLNTSTTTDDATSTSTPQASVDTATNTQTPAVTAPAPAVDPOONAVVSPQAV  
DTATPTVPTAATTPDGAAPLPTDQAGVTTTAADPNALVMNFTADCHLEVDATGSKLF  
SGWQRKQGNLNTGQAPYKLIKIGAPAAVQIQGKPVLSRIFRTINQVARTLNAEOS  
PAQ"  
complement (2534. .3688)  
/gene="yfgB"  
/locus\_tag="S2735"  
complement (2534. .3688)  
/gene="yfgB"  
/locus\_tag="S2735"  
/note="residues 1 to 384 of 384 are 99.47 pct identical to  
residues 1 to 384 of 384 from Escherichia coli K-12 :  
B2517"  
/codon\_start=1  
/transl\_table=1  
/product="hypothetical protein"  
/protein\_id="AAPI7890.1"  
/db\_xref="GI:30042165"  
/translation="MSQIVTPENVVTRTKDGKINLIDNROOMEFFKDLGKPTRADO  
VMKMYHYPCDNPEMTDINKVLEKGLKEVAETAPAPVVEQSSDGTITKAIAYGQ  
RVEVYIPEDRATLCVSQVGCALBECKFCSTAOQNRNIRVSEITIGQWRKAKTVG  
AAKTGQRPITNVVMGEGEPLNLNVVPMAMEIMLDFGGLSKRVTLVGRKATVGA  
LDKIDMDIVALAISLAENDEIDELVPIINKKNIITFLAAVARYYLEKSNANGVAT  
IEYVLDHVDNGTEHAQLAELLKDTPTCKINLIPWNPFPDAPVGRSSNSRIIDFSKVL

MSVGFITVIRKTRGDIDDAACQLAGVIDRTRKTLRKMOGEAIDIKAV"  
complement (3838. .4269)  
/gene="ndk"  
/locus\_tag="S2736"  
complement (3838. .4269)  
/gene="ndk"  
/locus\_tag="S2736"  
/function="enzyme; Purine ribonucleotide biosynthesis"  
/note="residues 1 to 143 of 143 are 100.00 pct identical  
to residues 1 to 143 of 143 from Escherichia coli K-12 :  
B2518"  
/codon\_start=1  
/transl\_table=1  
/product="nucleoside diphosphate kinase"  
/protein\_id="AAPI7891.1"  
/db\_xref="GI:30042166"  
/translation="MAIERTFSLIKRNAVAKVIGNIFARFEAGKIVGTQMLHLTV  
EQARKFYASHDKPFPDGLVERFTSGPIVSVLEGNAVQRHRLDGLATNPANALAGT  
LRADYADSLTENGTHGSDSVESAAREIAVFPGEVCPRTT"  
complement (4418. .6517)  
/gene="pbpc"  
/locus\_tag="S2737"  
/note="truncated"  
/pseudo  
6514. .7961  
/note="insertion element"  
/insertion\_seq="ISSf14"  
complement (6534. .7730)  
/locus\_tag="S2738"  
complement (6534. .7730)  
/locus\_tag="S2738"  
/function="IS, phage, Tn; Transposon-related functions"  
/note="residues 1 to 398 of 398 are 90.95 pct identical to  
residues 1 to 398 of 398 from GenBank :  
>db|AKI8438.1|AF348706.127 (AF348706) putative  
transposase for IS110 [Shigella flexneri]"  
/codon\_start=1  
/transl\_table=1  
/product="ISSf14 orf"  
/protein\_id="AAPI7892.1"  
/db\_xref="GI:30042167"  
/translation="MTSSDYBSVQVIGVDVGDTHHAVINRSGKRLFDKALPNDE  
NKRLSLIDLKQHQIILVDDQPAITGALPVAVAREGVYGLPGLAMRILADLHAG  
BAKTQVDAALITABARTLPHALRTIKLADQELASMCGFDDDLAAGTOGNSRIR  
GHLTYQHPAPBVRVGRPLHPAVALDQVYSPSEKIASIGKKLAAQLOKLAELDKR  
LADIAQALAEQTVVGGTNAAVVLPRLADITIRKQDEVALVEQVLAHPIYP  
VITSMGVAVRTAARLLTEVACRAFAAHLAAVAGLAVPTRSGSSIRGEHPSRGN  
KALKRALPLFSAPAAALRDPISRAYYTRKMSQGRHQNALALARRCDVLFAMWRDGT  
YFPQGS"  
7960. .8442  
/gene="ssea"  
/locus\_tag="S2740"  
/note="truncated"  
/pseudo  
complement (9261. .10037)  
/gene="sseaB"  
/locus\_tag="S2741"  
complement (9261. .10037)  
/gene="sseaB"  
/locus\_tag="S2741"  
/function="phenotype; Not classified"  
/note="residues 1 to 258 of 258 are 94.96 pct identical to  
residues 4 to 261 of 261 from Escherichia coli K-12 :  
B2522"  
/codon\_start=1  
/transl\_table=1  
/product="enhancer of serine sensitivity"  
/protein\_id="AAPI7893.1"  
/db\_xref="GI:30042168"  
/translation="MSETKLELDLEKATERRAPRFRITLSESTWVPGTAAGE  
AVEDSADLQDHEKEDGTSVLPFTSLDALQAVDEQAFVMPVPRITLPEMLTGLT  
FLNAKLPTEKEFNPRESILLLIEGNSLSOELLIEGSLISEVAEPAPQMDLSLT  
LFKTIKPVRAAFICSIKENDEAQPNLLIGIEADGDIIEIIQATGSVATDTLPDDEPID

CDS  
 locus\_tag="SF2561"  
 complement(6515..7642)  
 /gene="gcpE"  
 /locus\_tag="SF2561"  
 /note="Residues 1 to 372 of 375 are 99 pct identical to residues 1 to 372 of a 372 aa protein from Escherichia coli O157:H7 EDJ933 ref: NP\_289068.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="orf, conserved hypothetical protein"  
 /protein\_id="AA04061.1"  
 /db\_xref="GI:24052935"  
 /translation="MHNQAPIORRKRSTRIVGNVPIGDGAPAVOSMTNRTTVDVATVNOIKALERVGADIVRVSVPTMDAAEAFKIKORVNVPLVADIHFDYRIALKYAEVDCRINGNIGNEIRIMVDCARDKNIPIRIGNAGSLEKDIQEKGEPTPOLAESAMHVDLRLNDFDPKVSADVPVLAESVRLKQIDOPHLGITTEGASGAVKSAIGLGLSEGIQDTLRVSLADVPPEEIKVGDILKSLRISRGINFACPTCSROEFVIGTVALSEQRLEDITIPMVSIIIGVNVGEGALVSLGVTGKNKSLYEDGV RKDRLNNDMDIQLERIRAKASQLEARRIDVOVKIIT"  
 complement(7669..8682)  
 /gene="yfgA"  
 /locus\_tag="SF2562"  
 complement(7669..8682)  
 /gene="yfgA"  
 /locus\_tag="SF2562"  
 /note="Residues 1 to 337 of 337 are 99 pct identical to residues 1 to 337 of a 337 aa protein from Escherichia coli O157:H7 ref: NP\_311405.1"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative membrane protein"  
 /protein\_id="AA04062.1"  
 /db\_xref="GI:24052936"  
 /translation="NMTEATHDNEALTTGARLRNAREQIGLSQVAVERLCLKRSTVRDIEDAPADLASTPLRGITRSYARVHLPSEELPGEKQAPLRKAVAPWQSFSGKRRKRDGMFTWLVLFVVLGSGAMWQDKAQOEIITMADPOSSAESNSBOGQVNPSTTTPATPTSPASVDITATQTPAVTAPAPADPOONAVSPSOANVDTNATPVTATTPDGAAPLPDQAVTTPADPNALVMFTDCMLVTDATGKXLF

Query Match 98.24; Score 1098.4; DB 1; Length 10225;  
 Best Local Similarity 99.04; Pred. No. 1,1e-260;  
 Matches 1105; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGCATACACAGGCTCCATTCAACGAGAAATCAACAGCTATTTCAGTTGGGAATGTG 60  
 Db 7642 ATGCATACACAGGCTCCATTCAACGAGAAATCAACAGCTATTTCAGTTGGGAATGTG 7583  
 QY 61 CCGATTGGCGATGGTGTCCCATCGCCGATCAGTCCATGACCAATACGCGTACGACAGAC 120  
 Db 7582 CCGATTGGCGATGGTGTCCCATCGCCGATCAGTCCATGACCAATACGCGTACGACAGAT 7523  
 QY 121 GTGGAACACGCTCAATCAATCAAGCGCGCTGGAACGCGTGGCGGTGATATCGTCCT 180  
 Db 7522 GTGGAACACGCTCAATCAATCAAGCGCGCTGGAACGCGTGGCGGTGATATCGTCCT 7463  
 QY 181 GTATCGGTACGACGATGACGGGCGGAGAGCGTTCAACTCAATCAAGACGAGTTAAC 240  
 Db 7462 GTATCGGTACGACGATGACGGGCGGAGAGCGTTCAACTCAATCAAGACGAGTTAAC 7403  
 QY 241 GTGCCGCTGGTGGCTGACATCACTTCGATTCGCAATTCGCGTGAAGTAGGCAATAC 300  
 Db 7402 GTGCCGCTGGTGGCTGACATCACTTCGATTCGCAATTCGCGTGAAGTAGGCAATAC 7343  
 QY 301 GGGCTGATTTGTCTGGCTATTAAACCTGGCAATATCGGTATAGAGAGCTATTCCGATG 360  
 Db 7342 GGGCTGATTTGTCTGGCTATTAAACCTGGCAATATCGGTATAGAGAGCTATTCCGATG 7283  
 QY 361 GTGTGATCCTGTGCGCGCGATTAATAACATCCGATCCGTTTGGCGTTAAGCGCGATCG 420  
 Db 7282 GTGTGATCCTGTGCGCGCGATTAATAACATCCGATCCGTTTGGCGTTAAGCGCGATCG 7223  
 QY 421 CTGGAATAAAGATCTGCAAGAAAGATATGCGAACCGACCGGACGCGTGTGTAATCT 480

Db 7222 CTGGAATAAAGATCTGCAAGAAAGATATGCGAACCGACCGGACGCGTGTGTAATCC 7163  
 QY 481 GCCATGCGTATGTTGATCATCTCGATCGCGCTGAACCTTCGATCATGTTCAAGTCAGCGTG 540  
 Db 7162 GCCATGCGTATGTTGATCATCTCGATCGCGCTGAACCTTCGATCATGTTCAAGTCAGCGTG 7103  
 QY 541 AAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATCGTTTGTGCGCAAAACAGATCGAT 600  
 Db 7102 AAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATCGTTTGTGCGCAAAACAGATCGAT 7043  
 QY 601 CAGCGTTCGATCTGGGATCACGAAAGCGGTGTGGCGGACGCGGCGATTAATCC 660  
 Db 7042 CAGCGTTCGATCTGGGATCACGAAAGCGGTGTGGCGGACGCGGCGATTAATCC 6983  
 QY 661 GCCATGTTTATGATCTGCTGCTGCTGCTGAAGCATGGGGAACGCTGGCGGATGCGTG 720  
 Db 6982 GCCATGTTTATGATCTGCTGCTGCTGCTGAAGCATGGGGAACGCTGGCGGATGCGTG 6923  
 QY 721 GCGGCGATCCGCTCGAAGAGATCAAGTTCGATATTTTGAATTCGCTGCTATC 780  
 Db 6922 GCGGCGATCCGCTCGAAGAGATCAAGTTCGATATTTTGAATTCGCTGCTATC 6863  
 QY 781 CGTTCCGAGGAGATCACTTCATCGCTCGCGACCTGTTCCGCTGACGAAATTTGATGT 840  
 Db 6862 CGTTCCGAGGAGATCACTTCATCGCTCGCGACCTGTTCCGCTGACGAAATTTGATGT 6803  
 QY 841 ATCGATACGTTTAAAGCGCTGAGCAAGCGCTGGAAGATATCACTTCGATGAGCGTT 900  
 Db 6802 ATCGATACGTTTAAAGCGCTGAGCAAGCGCTGGAAGATATCACTTCGATGAGCGTT 6743  
 QY 901 TCGATATCGGCTGCTGCTGCTGATGAGCCGACGAGTGAAGCGCTGTTCTACACTCGGCGTC 960  
 Db 6742 TCGATATCGGCTGCTGCTGCTGATGAGCCGACGAGTGAAGCGCTGTTCTACACTCGGCGTC 6683  
 QY 961 ACCGCGGCAACAAAGAAAGCGGCTCTATGAAGATGCGCTGCGCAAGACGCTTGAGAC 1020  
 Db 6682 ACCGCGGCAACAAAGAAAGCGGCTCTATGAAGATGCGCTGCGCAAGACGCTTGAGAC 6623  
 QY 1021 AACAAAGATATGATGACACGAGTGAAGACGAGTTCGCGGAACGAGTCAAGTCAAGTGCAGC 1080  
 Db 6622 AACAAAGATATGATGACACGAGTGAAGACGAGTTCGCGGAACGAGTCAAGTCAAGTGCAGC 6563  
 QY 1081 GAAGCGCTGCAATGACCTTCAGCAGTTGAAAAA 1116  
 Db 6562 GAAGCGCTGCAATGACCTTCAGCAGTTGAAAAA 6527

RESULT 8  
 AB016987/c  
 LOCUS  
 DEFINITION  
 Shigella flexneri 2a str. 2457T section 10 of 16 of the complete genome.  
 ACCESSION  
 AB016987 AB014073  
 VERSION  
 AB016987.1 GI:30042162  
 SOURCE  
 ORGANISM  
 Shigella flexneri 2a str. 2457T  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.  
 REFERENCE  
 1 (bases 1 to 290380)  
 Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R.  
 Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T  
 Infect. Immun. 71 (5), 2775-2786 (2003)  
 JOURNAL  
 PUBMED  
 12704152  
 2 (bases 1 to 290380)  
 Wei,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W., Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A., Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S., Schwartz,D.C. and Blattner,F.R.

TITLE Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157

JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)

PUBMED 12384590

REFERENCE (bases 1 to 10225)

AUTHORS Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, F., Zhang, X.B., Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.O., Wang, Y., Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.

TITLE Direct Submission

JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China

FEATURES

source

1. 10225

/location/Qualifiers

/organism="Shigella flexneri 2a str. 301"

/mol\_type="genomic DNA"

/strain="301"

/serotype="2a"

/db\_xref="taxon:198214"

148. 1440

/gene="xsea"

/locus\_tag="SF2555"

148. 1440

/gene="xsea"

/locus\_tag="SF2555"

/note="Residues 1 to 430 of 430 are 97 pct identical to residues 27 to 456 of a 456 aa protein from *Escherichia coli* K12 ref: NP\_417004.1"

/codon\_start=1

/transl\_table=11

/product="exonuclease VII, large subunit"

/protein\_id="AA04055.1"

/db\_xref="GI:24052929"

/translation="MGQWVTSGEISNFTOPASGHWYPTLKDDTAQVRCAMFRNSNRVTPPHOGQOVLVRANITLIEPRGDIIVSMOPAGGLQOKEOLKALQACLEDQOYKPLSPAHCVGTSTKGAHLHLIVLKRDPSPILVITPTSVGGDPAQGVIRAIENORNECVLIVRGSGGSLIEDMSFNDERVARATPASRIPTVSAGCHETDVIADFVADRAPTPSAAEVSVRNOQELIRQVSRORLEMDYLLARTIRFPQIHHRLOOQHPLRLAQQTMLERLOKMSFALENQLKRAQGOQRLTORLNOQMPQKHHTOTRIQOLEVRLAEILRAQLSATRRFENATHLAEVSPSLTARGYSTTATGDNVLYKKVQKGTGEMLTTRLEDGVIIESEYKNIQPVKSKKKVH"

complement(1437..1688)

gene

/gene="yfgu"

/locus\_tag="SF2556"

complement(1437..1688)

/gene="yfgu"

/locus\_tag="SF2556"

/note="Residues 1 to 83 of 83 are 97 pct identical to residues 1 to 83 of a 83 aa protein from *Escherichia coli* K12 ref: NP\_417005.1"

/codon\_start=1

/transl\_table=11

/product="orf, conserved hypothetical protein"

/protein\_id="AA04056.1"

/db\_xref="GI:24052930"

/translation="MNVGMATGGINHMLHCPKQHVLDQNGHARCPSCGEFEMKALCPDQHQLQVLAKGAVDYFCOHGLISKVVEFLA"

complement(1721..3332)

gene

/gene="yfgk"

/locus\_tag="SF2557"

complement(1721..3332)

/gene="yfgk"

/locus\_tag="SF2557"

/note="Residues 1 to 503 of 503 are 99 pct identical to residues 1 to 503 of a 503 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP\_289064.1"

/codon\_start=1

/transl\_table=11

/product="putative GTP-binding factor"

/protein\_id="AA04057.1"

/db\_xref="GI:24052931"

gene

/translation="NRCLMTIYKNEALNMPVVALVGRPNVGKSTLFNRLTRDALVADPPGLRDRKRGARLEIEBEFCIDTGGDGEDGETMAESLSLAIEADVLPMVDPACGLMPADPAIAXLIRSRKPTPLVNAQTGLDPODAVDFPSGLDEIYVIAASHGNGVLSLHVLILPMMEDLAPQEVDEDELYNAQFPAENGEEBDDDPQSPILKALIVGRPNVKSSTLTKNLIGERVVYVDMGTDRSDSYITIMERDGRVVLITAGVKRGITDAVEKESVITKLOAIEDANVMVLVIDAREGISDQSLSLGFLINSGRSLVIVNKNMGDSQEVKEQVETLDRGLIDPFAVHFIASLGSQVGNLFESVRAVYOSTERVGTSMRLRTMTMAVEDHQPVLVGRVRYKLYAHAGVYNPPIVTHGQVRLDPSYKRYLNNYPRKSLDWNQSPTRIOFKEGENYANKRNTLTPTQMRKKRLMKHLMKK"

complement(3311..4489)

gene

/gene="yfgl"

/locus\_tag="SF2558"

complement(3311..4489)

/gene="yfgl"

/locus\_tag="SF2558"

/note="Residues 1 to 392 of 392 are 99 pct identical to residues 1 to 392 of a 392 aa protein from *Escherichia coli* K12 ref: NP\_417007.1"

/codon\_start=1

/transl\_table=11

/product="putative dehydrogenase"

/protein\_id="AA04058.1"

/db\_xref="GI:24052932"

/translation="MQLRLILPGLLSVTLSGCSLFNSEEDVYKNSPLPTVENQPTPTAWSTVSIGINIFYSNLHPALADNVVAAADRAGLVAKLADDDKEIWSVLSAEDGWFSKEPALLSGGVTSGGHVYIGSEKAOVYALNTSGTAMQTKVAGALSPPVSDGLVLIHNSQLOALNADADAVKVTYVLMPSLSIRESAPTATFGAIVGQNGRVSALVMEQGMIMQORISQATGSTIEDRLSDVDTTPVYVNGVFALATNGNTALDLRSGQIMVKRELGSVNFIVDGNRIYVLDQDRWALITDGGVILMQSDILHLLTSPVLYN GNLVVDSESGYLMWINEDEGRFVAQKVDSSGFPQEPVADKLLIQANDGVYSITR"

complement(4500..5120)

gene

/gene="yfgm"

/locus\_tag="SF2559"

complement(4500..5120)

/gene="yfgm"

/locus\_tag="SF2559"

/note="Residues 1 to 206 of 206 are 98 pct identical to residues 1 to 206 of a 206 aa protein from *Escherichia coli* K12 ref: NP\_417008.1"

/codon\_start=1

/transl\_table=11

/product="orf, conserved hypothetical protein"

/protein\_id="AA04059.1"

/db\_xref="GI:24052933"

/translation="MEIYENDDQVEAVKRFPAENGKALAVGSLGVALTCKRYNNSHOVDSARSASLAVYONAVTAVSBGKPSIPAAEKFAENKNTYGALSLSLAQOFPVKNELKAAAOLOQGLADTSDENLRKAVINLRARVOVLOKQADALKTIDTITKGGWMAIVADLRGALLSKGKGQARSAMEKGVSDVTPALSEMMQKINNLSTI"

complement(5138..6412)

gene

/gene="hiss"

/locus\_tag="SF2560"

complement(5138..6412)

/gene="hiss"

/locus\_tag="SF2560"

/note="Residues 1 to 424 of 424 are 99 pct identical to residues 1 to 424 of a 424 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP\_289067.1"

/codon\_start=1

/transl\_table=11

/product="histidine tRNA synthetase"

/protein\_id="AA04060.1"

/db\_xref="GI:24052934"

/translation="MAKNIQAIRGNADYIPGETAIVMORIEGLKAVLVSGYSEILPIVEQTPLFRAIGEVTDVVEKMYTFEDNNGSLTRPAGTGCVAAGIEHGLVIOEORLWYIGPMFRHRPQKGRYQFHOQCEVFGDIDAEILMLVARWRALGFIETHLEINSIGSLBARAPDALVAFLQHKEXLDEDCKRMVTNPLVLDSKNEVQALLNDAPRLGVLDEBSREHPDAGLCKLESAGIAYTVNQRVLRGLDYNRTVFENVTSLGSGCTVACAGRGYGLVEOLGGRATPAVAGKAGLERVILLVQAVNPEFKADPDVDTLVASGADTQSAAMALERLDELGLVYKLMTHGGGSPFKQPARADKMARAAVVLGSEBVANGTAVVVDLRGEGQTAVADSVAAHLRTLLG"

complement(6515..7642)

gene

/gene="gcpB"



REFERENCE	TITLE
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.	
1 (bases 1 to 13176)	
Riley,M., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Maynew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.	
The complete genome sequence of Escherichia coli K-12 Science. 277 (5311), 1453-1474 (1997)	
2 (bases 1 to 13176)	
Blattner,F.R.	
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecol@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
3 (bases 1 to 13176)	
Blattner,F.R.	
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA. Email: ecol@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459	
4 (bases 1 to 13176)	
Plunkett,G. III.	
Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA	
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH Grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GenMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Beryn. A public version of the database is accessible ( <a href="http://cgsc.biology.yale.edu">http://cgsc.biology.yale.edu</a> ). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site ( <a href="http://www.genetics.wisc.edu">http://www/genetics.wisc.edu</a> ). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.	
location/Qualifiers	
1. 13176	
/organism="Escherichia coli K12"	
/mol_type="genomic DNA"	
/strain="K12"	
/sub_strain="MG1655"	
/db_xref="taxon:83333"	
/complement(49.. 81)	
/note="factor Sigma70; promoter hns; documented +1 at 2638662"	
gene	
complement(86.. 1204)	
/gene="ygcB"	
/note="synonym: b2515"	
complement(86.. 1204)	
/gene="ygcB"	
/function="orf; Unknown"	
CDS	
/note="#372; 100 pct identical to GCPE_ECOLI SW: P27433"	
/codon_start=1	
/transl_table=11	
/product="orf, hypothetical protein"	
/protein_id="AAC75568.1"	
/db_xref="GI:1788863"	
/translation="MHNOAPIORRSTRIYGVNPIGDGAPIAVOSMTTRITDVATVNOIKALRRVCADIVRVSPPTMDAAEAFKLIOOVNPVLADHEDVRILAKAEVDCLRINPENGIRNEIRNVUDCARDKNIPRTIGNASSLKDQEKGEPTPOALLSARRHVHLDRNLFDQFKVSASVDPLAVERSLAKOIQDPLHGITEGGARSGVAKSAIGIGLLSEGIIDTIRLSLDPAVEELKVGFDDILKSIRIRSRGINFLACPSCSQBEDVIGTVNALDEQRLEDIIITPMDSIIICCVNGCEALVSTLTGTGNKKSGJEDGVRCRDNDNDMIIDQLEARIRAKASQLEARRRIDVOQVER"	
complement(1231.. 2244)	
/gene="yfgA"	
/note="synonym: b2516"	
complement(1231.. 2244)	
/gene="yfgA"	
/function="putative membrane; Not classified"	
/note="#337; 100 pct identical to YFGA_ECOLI SW: P27434"	
/codon_start=1	
/transl_table=11	
/product="putative membrane protein"	
/protein_id="AAC75569.1"	
/db_xref="GI:1788864"	
/translation="WNTBATHDONBALTTGARLRNAREFQSLGQAVERLCIKSVTVRLIEEDKAPADLASTFTLGRTYSTARLVHPLEBLPGLEKQAPRAKAPMOSBELGRRKRKGDMWTFTFWLFPVIGLSGMMWQDKAQOEITTMADSSLELSNSQGSVPILNSTTTPDTATSTPASVDTPTATNOTPAVAPAADPCQNAVSSPSQALVDPAAEPAPTAATPPGAAPLPDAGVTVTPADNAPVALVMNFETADGTCGLFSGMKRKDNINLTGAPPKLKIGAPNAVQIQGKPVDLSFRITNQVARLTINAEGSPAO"	
complement(1288.. 1316)	
/gene="yfgA"	
/note="factor Sigma70; predicted +1 start at 2639901"	
complement(1360.. 1389)	
/gene="yfgA"	
/note="factor Sigma70; predicted +1 start at 2639973"	
complement(2529.. 3683)	
/gene="yfgB"	
/note="synonym: b2517"	
complement(2529.. 3683)	
/gene="yfgB"	
/function="orf; Unknown"	
/note="#184; 100 pct identical to YFGB_ECOLI SW: P36979"	
/codon_start=1	
/transl_table=11	
/product="orf, hypothetical protein"	
/protein_id="AAC75570.1"	
/db_xref="GI:1788865"	
/translation="MSBOJLYPVENVTKDGKINLLDNRQOREFFKDLGKPRPADQWKMMHYCCDNFDEMDFINKYLKKEVEIAPEVEBSQSDGTIKWALVAGQOEVREYIVIPEDDPAATCVSSOGCALKECFSTAOGFNRLRVEIIGOWRAKVIQALAKVTMGORPITNVAMGMGEPFINNNVPAIFDGLSKRSKVVTSISGVPA LRKGMIDMALAISLARPNBEIRDELIVPKKNMIEFTLAAYRYLEKSNANGRPV LIEVWLIDHWNDSTERHAQDLIELKTDPCKNIKLIIPWNPFGAPYGRSSNSRIDRESKVL MSYGFTTIWRKTRGDIDIAACGOLAADVIDTRKTLRRMQGEAIDIKAV"	
complement(3778.. 3806)	
/note="factor Sigma70; predicted +1 start at 2642391"	
complement(3786.. 3814)	
/note="factor Sigma70; predicted +1 start at 2642399"	
complement(3833.. 4264)	
/gene="ndk"	
/note="synonym: b2518"	
complement(3833.. 4264)	
/gene="ndk"	
/EC_number="2.7.4.6"	
/function="enzyme; Purine ribonucleotide biosynthesis"	
/note="#143; 100 pct identical to NDK_ECOLI SW: P24233"	
/codon_start=1	
/transl_table=11	
/product="nucleoside diphosphate kinase"	
/protein_id="AAC75571.1"	



VERSION X64451.1 GI:41540  
KEYWORDS gcpE gene.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
REFERENCE Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
AUTHORS Enterobacteriaceae; Escherichia.  
TITLE 1 (bases 1 to 1697)  
JOURNAL Parker, J.  
Direct Submission  
Submitted (17-FEB-1992) J. Parker, Southern Illinois Univ, Dept of  
Microbiology, S I U, Carbondale IL 62901, USA  
REFERENCE 2 (bases 1 to 1697)  
AUTHORS Baker, J., Franklin, D.B. and Parker, J.  
TITLE Sequence and characterization of the gcpE gene of Escherichia coli  
JOURNAL FEMS Microbiol. Lett. 94, 175-180 (1992)  
COMMENT See also J01629 & M1843.  
FEATURES  
source Location/Qualifiers  
1..1697  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K12"  
/db\_xref="taxon:562"  
/sub\_clone="pSE411"  
/cell\_line="JF298"  
/clone\_id="lambda transducing phage dguab-2"  
<1..508  
/note="ORF"  
/codon\_start=2  
/transl\_table=1  
/protein\_id="CA45782.1"  
/db\_xref="GI:41541"  
/db\_xref="SWISS-PROT:P27434"  
/translation="TTTDPATTSPPASVVTATNTOTPAVTAAPAPAPQQAUVSP  
SCANVTPAPAPAPATTPGAALPTPDQAGVTPVADPAPALVNPFTADCELTXTD  
GKLFSGMQRKDMNLFGQAPYKLTGAPAAVQIQGKPVDSFRITNOVARLTL  
NAQSPAQ"  
535..1653  
/gene="gcpB"  
535..1653  
/gene="gcpB"  
/gene="gcpB"  
/codon\_start=1  
/transl\_table=1  
/protein\_id="CA45783.1"  
/db\_xref="GI:41542"  
/db\_xref="SWISS-PROT:P27433"  
/translation="MKNQAPIQRKRSRIYGVNVPIDGAPAPVQSMTRTDTVEAT  
VNOIKALERVADIVRVSPPTMAAEAFKLIKQVNPVLADIHFDRIALKVAEYGV  
DCIRINPQNGNERIRMVVDCADKNIPIRIGNAGSLBKDLQEKYGEPTPALLES  
AMRHVDLDRLEDFQFVSVKASDVFLAVSYRLAKOIQDPLHGLITTEAGARGAV  
KSAIGLILSEIGDILRVSLAADPVEBIRKVGFDLKSIRSRGINFTACPTCSRO  
EPVIGCTVNALEORLEDTITPMVSIIGCVNPGPEALVSTLGVTGSKKSGLYEDCV  
RKRLDNDNDMDIDEARIRAKASQLDARHIDVOQVEK"

BASE COUNT 417 a 446 c 469 g 365 t  
ORIGIN

Query Match 100.0%; Score 1119; DB 1; Length 1697;  
Best Local Similarity 100.0%; Pred. No. 8.7e-266;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACCAAGGCTCCATTCACGTAAGAAATCAACAGTATTTACGTTGGGAATGTG 60  
DB 535 ATGCATACCAAGGCTCCATTCACGTAAGAAATCAACAGTATTTACGTTGGGAATGTG 594  
QY 61 CCGATTGGCATGTGCTCCCATGCGCTACAGTCCATGACCAATACGCGTACGACAGAC 120  
DB 595 CCGATTGGCATGTGCTCCCATGCGCTACAGTCCATGACCAATACGCGTACGACAGAC 654  
QY 121 GTGAGCAACGCGTCATCAATCAAGCGCTGGAAGCGCTGCGCTGATATCTGCTCGCT 180  
DB 655 GTGAGCAACGCGTCATCAATCAAGCGCTGGAAGCGCTGCGCTGATATCTGCTCGCT 714  
QY 181 GTATCCGTACCGACGATGACGCGGACGAAGCGTTCAAAGTCAATCAACAGCAGTTAAC 240  
|||||

DB 715 GTATCCGTACCGACGATGACGCGGACGAAGCGTTCAAAGTCAATCAACAGCAGTTAAC 774  
QY 241 GTGCGCGTGGTGGCTGACATCCATTGCAATTCGATTCGCTGGTGAAGTGGGAATAC 300  
DB 775 GTGCGCGTGGTGGCTGACATCCATTGCAATTCGATTCGCTGGTGAAGTGGGAATAC 834  
QY 301 GGCCTGCATTTGCTGCTATTTAAACCTGCAATATCGTAAATGAGACGCTATTCGATG 360  
DB 835 GGCCTGCATTTGCTGCTATTTAAACCTGCAATATCGTAAATGAGACGCTATTCGATG 894  
QY 361 GTGCTTACCTGTGCGCGCATTAATAACATTCGATTCGATTCGCTTAAACCGCGATCG 420  
DB 895 GTGCTTACCTGTGCGCGCATTAATAACATTCGATTCGATTCGCTTAAACCGCGATCG 954  
QY 421 CTGAAAAAATGTCGACGAAGAAAGTATGGGAAACGACGCGGAGCGTGTGGAATCT 480  
DB 955 CTGAAAAAATGTCGACGAAGAAAGTATGGGAAACGACGCGGAGCGTGTGGAATCT 1014  
QY 481 GCCATGCGTATGTTGATCATCTCGATCGCCTGAACCTTCGATCAATCAAGTCAGCGTG 540  
DB 1015 GCCATGCGTATGTTGATCATCTCGATCGCCTGAACCTTCGATCAATCAAGTCAGCGTG 1074  
QY 541 AAGCGTCTGACGCTCTTCTCGCTGTTGAGTCTTATTCGTTTCTGCGCAAAACAGATCAT 600  
DB 1075 AAGCGTCTGACGCTCTTCTCGCTGTTGAGTCTTATTCGTTTCTGCGCAAAACAGATCAT 1134  
QY 601 CACGCGTGCATCTGGGGATCAACGAAGCCGGTGGTGGCGGACGGGGAGTAAATCC 660  
DB 1135 CACGCGTGCATCTGGGGATCAACGAAGCCGGTGGTGGCGGACGGGGAGTAAATCC 1194  
QY 661 GCCATTGTTTAAAGTCTGCTGCTGCTGTAAGGATCGGACACAGCTGCGCATTCGCTG 720  
DB 1195 GCCATTGTTTAAAGTCTGCTGCTGCTGTAAGGATCGGACACAGCTGCGCATTCGCTG 1254  
QY 721 GCGCGCGATCGGTCGAAGATCAAAAGTCGTTTGCATATTTTGAATTCGTCGCTATC 780  
DB 1255 GCGCGCGATCGGTCGAAGATCAAAAGTCGTTTGCATATTTTGAATTCGTCGCTATC 1314  
QY 781 CGTTCCGAGGAGATCAACTTCATCGCTCGCCGACCTGTTCCGGTACGGAATTTGATGTT 840  
DB 1315 CGTTCCGAGGAGATCAACTTCATCGCTCGCCGACCTGTTCCGGTACGGAATTTGATGTT 1374  
QY 841 ATCGGTACGGTTTAAAGCTGCTGCTGCTGACCAAGCCTGGAAGATTCATCACTCCGATGACGTT 900  
DB 1375 ATCGGTACGGTTTAAAGCTGCTGCTGCTGACCAAGCCTGGAAGATTCATCACTCCGATGACGTT 1434  
QY 901 TCGATTATCGGCTGCTGCTGCTGAAATGGCCAGGTGAGCGCTGCTTCTACACTCGGCGTC 960  
DB 1435 TCGATTATCGGCTGCTGCTGCTGAAATGGCCAGGTGAGCGCTGCTTCTACACTCGGCGTC 1494  
QY 961 ACCGGCGGCAACGAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAAAGCCTTCGAC 1020  
DB 1495 ACCGGCGGCAACGAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAAAGCCTTCGAC 1554  
QY 1021 AACCAAGATATGATGACCAAGCCTGGAAGACGCGATTCGTGCGAAAGCAGTCACTGAC 1080  
DB 1555 AACCAAGATATGATGACCAAGCCTGGAAGACGCGATTCGTGCGAAAGCAGTCACTGAC 1614  
QY 1081 GAAGCGCGTGCATTTGACGTTGACGAGGTTGAATAATA 1119  
DB 1615 GAAGCGCGTGCATTTGACGTTGACGAGGTTGAATAATA 1653  
RESULT 6  
AE000338/c 13176 bp DNA linear BCT 01-DEC-2000  
LOCUS Escherichia coli K12 MG1655 section 228 of 400 of the complete  
DEFINITION  
ACCESSION AE000338 U00096  
VERSION AE000338.1 GI:1788862  
KEYWORDS  
ORGANISM Escherichia coli K12



QY 841 ATGGTACGGTTAAACGGCTGAGACCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900  
Db 841 ATGGTACGGTTAAACGGCTGAGACCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900  
QY 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTTCTACACTCGAGCTC 960  
Db 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTTCTACACTCGAGCTC 960  
QY 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020  
Db 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020  
QY 1021 AACACGATATGATTCGACCGAGCTGGAAGCAAGCATTTGTCGGAAGGCAATCTGAGTAC 1080  
Db 1021 AACACGATATGATTCGACCGAGCTGGAAGCAAGCATTTGTCGGAAGGCAATCTGAGTAC 1080  
QY 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119  
Db 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119

RESULT 4  
AX393943 1119 bp DNA linear PAT 23-MAR-2002  
LOCUS Sequence 3 from Patent WO0212478.  
DEFINITION AX393943  
VERSION AX393943.1 GI:19701905  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.

REFERENCE  
AUTHORS 1  
Boronat, A., Campos, N., Rodriguez-Concepcion, M., Rohmer, M.,  
Seeman, M., Valentin, H. E., Venkatesh, T. V. and Venkatesh, M.  
TITLE Methyl-d-erythritol phosphate pathway genes  
JOURNAL Patent: WO 0212478-A 3 14-FEB-2002;  
Monsanto Technology LLC (US)  
FEATURES  
source  
1. 1119  
Location/Qualifiers  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
1. 1119  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD29065.1"  
/transl\_table=1  
/db\_xref="GI:19701906"  
/translation="MHNQAPIQRKSTRIVGNVPIGDGAPLAVOSMTNTRTDEAT  
VNOIKALERVGADIVRVSPTMDAAEFKLIKQOVNPLVADHFDVRLKAKAEV  
DCIRINGENIERIMVVCARDKNIPIRIGVNASLEKDEKGEPTPOLLES  
AMRVHDLRLNFPQFVSVKASDVPLAVESYRLAKOIPHLGITTEAGRGSGV  
KSAIGTILSEIGDPLRVSILADPVEIKVGFIDIKSIRISRGINFIACPTCSRO  
EFVDITGVNLABQRLIEDITTPMDVSIIGCVINVEGELVSTLGVTSKKSGLYEDSV  
RKRLDNDNDIMIDLEARIKAKASQLEARRIDVQVEX"

BASE COUNT 272 a 279 c 317 g 251 t  
ORIGIN

Query Match 100.0%; Score 1119; DB 6; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 8.6e-266;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACACGAGCTCCATTCACAGTAAATCAACAGCTATTATTCGTTGGAGATGTG 60  
Db 1 ATGCATACACGAGCTCCATTCACAGTAAATCAACAGCTATTATTCGTTGGAGATGTG 60  
QY 61 CCGATTGGCGATGGTGTCTCCCATGCGGTACAGTTCATGACCAATACGCGTACAGAC 120  
Db 61 CCGATTGGCGATGGTGTCTCCCATGCGGTACAGTTCATGACCAATACGCGTACAGAC 120  
QY 121 GTGAGACACGCGTCAATCAAAATCAAGCGCTGAGAGCGGTGGCGGTATATGTCGCT 180

Db 121 GTGAGACACGCGTCAATCAAAATCAAGCGCTGAGAGCGGTGGCGGTATATGTCGCT 180  
QY 181 GTATCCGTACCGACATGAGACGCGGACAGAGCGTTCAACTATCAACAGAGTTAAC 240  
Db 181 GTATCCGTACCGACATGAGACGCGGACAGAGCGTTCAACTATCAACAGAGTTAAC 240  
QY 241 GTGCCCGCTGTGCTGACATCCACTTCGACTATTCGATTCGCTGAGAGTGAAGTACGGAATTC 300  
Db 241 GTGCCCGCTGTGCTGACATCCACTTCGACTATTCGATTCGCTGAGAGTGAAGTACGGAATTC 300  
QY 301 GGCGTGAATTCGCTGCTGATTAACCTGCAATATTCGTAATGAAAGGCTATTCAGATC 360  
Db 301 GGCGTGAATTCGCTGCTGATTAACCTGCAATATTCGTAATGAAAGGCTATTCAGATC 360  
QY 361 GTGCTGACTGTGCGCGCATTAATAATTCGATTCGATTCGCTGATTCGCTGATTCG 420  
Db 361 GTGCTGACTGTGCGCGCATTAATAATTCGATTCGATTCGCTGATTCGCTGATTCG 420  
QY 421 CTGGAATAAGATCTGCAAGAAAGTATGGGGAACGACCGCGGCGTTGCTGGAATCT 480  
Db 421 CTGGAATAAGATCTGCAAGAAAGTATGGGGAACGACCGCGGCGTTGCTGGAATCT 480  
QY 481 GCCATCGCTCATGCTGATCATCTCGATCCGCTGAACCTTCGATTCGATTCGATTCG 540  
Db 481 GCCATCGCTCATGCTGATCATCTCGATCCGCTGAACCTTCGATTCGATTCGATTCG 540  
QY 541 AAAGCGTCTGACGCTCTCTCTGCTGCTGATTCGATTCGCTGCTGCTGCTGCTGCTG 600  
Db 541 AAAGCGTCTGACGCTCTCTCTGCTGCTGATTCGATTCGCTGCTGCTGCTGCTGCTG 600  
QY 601 CAGCGCTTCATCTGCGGATCAACCGAAGCGGCTGCGCGGCGGCGGAGTAAATTC 660  
Db 601 CAGCGCTTCATCTGCGGATCAACCGAAGCGGCTGCGCGGCGGCGGAGTAAATTC 660  
QY 661 GCCATTGTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
Db 661 GCCATTGTTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720  
QY 721 GCGCGCGATCCGCTGCAAGAGATCAAAAGTTCGATTCGATTCGATTCGATTCG 780  
Db 721 GCGCGCGATCCGCTGCAAGAGATCAAAAGTTCGATTCGATTCGATTCGATTCG 780  
QY 781 CGTTCCGAGGATTCATTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
Db 781 CGTTCCGAGGATTCATTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840  
QY 841 ATCGTACGCTTAAACGCTGAGCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900  
Db 841 ATCGTACGCTTAAACGCTGAGCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900  
QY 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTTCTACACTCGAGCTC 960  
Db 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTTCTACACTCGAGCTC 960  
QY 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020  
Db 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020  
QY 1021 AACACGATATGATTCGACCGAGCTGGAAGCAAGCATTCGTCGGAAGGCAATCTGAGTAC 1080  
Db 1021 AACACGATATGATTCGACCGAGCTGGAAGCAAGCATTCGTCGGAAGGCAATCTGAGTAC 1080  
QY 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119  
Db 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119

RESULT 5  
ECGCP  
LOCUS ECGCP  
DEFINITION E.coli gcp gene.  
ACCESSION X64451 S43432  
1697 bp DNA linear BCT 08-DEC-1993

QY 421 CTGAAAAAGATCTGCAAGAAAAGATGCGGACCGACCGCGCGGCGTTCGGAATCT 480  
 DB 421 CTGAAAAAGATCTGCAAGAAAAGATGCGGACCGACCGCGCGGCGTTCGGAATCT 480  
 QY 481 GCCATGGCTCATGTTGATCTCTGATCGCTGAATTTCCATCTCAAGTCAAGCTG 540  
 DB 481 GCCATGGCTCATGTTGATCTCTGATCGCTGAATTTCCATCTCAAGTCAAGCTG 540  
 QY 541 AAGCGCTGACGCTCTCTCGCTGTTGAGCTTATGTTTGTGCGCAAAACAGATCAT 600  
 DB 541 AAGCGCTGACGCTCTCTCGCTGTTGAGCTTATGTTTGTGCGCAAAACAGATCAT 600  
 QY 601 CAGCGCTGATCTGGGATCAACGAAAGCGGTGTCGCGCGACGGGAGTAAATCC 660  
 DB 601 CAGCGCTGATCTGGGATCAACGAAAGCGGTGTCGCGCGACGGGAGTAAATCC 660  
 QY 661 GCCATTGGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGCGCTATCGCTG 720  
 DB 661 GCCATTGGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGCGCTATCGCTG 720  
 QY 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780  
 DB 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780  
 QY 781 CGTTCCGAGGATCAACTTCATCGCTCCGACCTGTCGCTCAGAAATTTGATGTT 840  
 DB 781 CGTTCCGAGGATCAACTTCATCGCTCCGACCTGTCGCTCAGAAATTTGATGTT 840  
 QY 841 ATGGGTTCGGTTAAACGCTGAGCAACGCTGGAAGATTCATCACTCCGATGAGCTT 900  
 DB 841 ATGGGTTCGGTTAAACGCTGAGCAACGCTGGAAGATTCATCACTCCGATGAGCTT 900  
 QY 901 TCGATTATCGGCTGCGTGAATGCGCAAGTGAAGGCGGTGTTTCACTCGGCTGTC 960  
 DB 901 TCGATTATCGGCTGCGTGAATGCGCAAGTGAAGGCGGTGTTTCACTCGGCTGTC 960  
 QY 961 ACCGCGCGCAACAGAAAAGCGGCTCTATGAAAGTGGCTGCGCAAAAGCCGCTGAC 1020  
 DB 961 ACCGCGCGCAACAGAAAAGCGGCTCTATGAAAGTGGCTGCGCAAAAGCCGCTGAC 1020  
 QY 1021 AACACGATATGATCGACGCTGGAAGCAACGCTTGTGCGAAAGCCAGTCACTGAGC 1080  
 DB 1021 AACACGATATGATCGACGCTGGAAGCAACGCTTGTGCGAAAGCCAGTCACTGAGC 1080  
 QY 1081 GAAGCGCGTGAATTGAGCTTCAAGAGTTGAAAAATTA 1119  
 DB 1081 GAAGCGCGTGAATTGAGCTTCAAGAGTTGAAAAATTA 1119

RESULT 3  
 AX050487 1119 bp DNA linear PART 12-JAN-2001  
 LOCUS AX050487  
 DEFINITION Sequence 1 from Patent WO0072022.  
 ACCESSION AX050487  
 VERSION AX050487.1 GI:12226691  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 Enterobacteriaceae; Escherichia.

REFERENCE  
 AUTHORS Jomaa, H.  
 TITLE Use of genes of the deoxy-d-xylose phosphate biosynthetic pathway  
 JOURNAL for altering the concentration of Isoprenoid  
 Patent: WO 0072022-A 1 30-NOV-2000;  
 Jomaa, Hasaan (DE)

FEATURES  
 source  
 1. 1119  
 location/Qualifiers  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 1. 1119  
 /note="unnamed protein product"

/codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC1801.1"  
 /db\_xref="GI:12226692"  
 /translation="MHNQAPLQRRKSTRLYGVNPVIGDAPLPAVQSMNTRTTVDAT  
 VNOIKALERVADIVRVSPTMDAAEFLIKQVNPVADLHFDYRALKVAEYGV  
 DCLRIHPNIGNEERIRNVDCAKDNIRIRGVNAGSELEKIDOEYGEPTPALLES  
 ARNHYDNLNFDQKVSADVLAESYRLAKQIDOPHLGITEAGASGAV  
 KSAIGGLLSRIGDILRLVLAADVEEIKVGFDLKSLRISRSRGINFLACPTGSRQ  
 EFDVCTNALRQRELDITTPMDVSIIGCVNRPBALYSTLGVTGNKKSGLYEDGV  
 RDRLENNMIDLEARIKASQLEBARRIDVQVER"

BASE COUNT 272 a 279 c 317 g 251 t

ORIGIN

Query Match 100.0%; Score 1119; DB 6; Length 1119;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-266;  
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTACGTTGGAAATG 60  
 DB 1 ATGCATACAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTACGTTGGAAATG 60  
 QY 61 CCGATTGGCGATGCTCTCCATCGCTGACATCATGATCAACATACCGGTACAGAC 120  
 DB 61 CCGATTGGCGATGCTCTCCATCGCTGACATCATGATCAACATACCGGTACAGAC 120  
 QY 121 GTGAGCAACGGTCAATCAAAATCAAGGGCGTGGACGGCTGGCGCTGATATCGTCC 180  
 DB 121 GTGAGCAACGGTCAATCAAAATCAAGGGCGTGGACGGCTGGCGCTGATATCGTCC 180  
 QY 181 GTATCCGTACCGACGATGACGCGGAGAGAGCTTCAAACTCATCAACAGCAGTTAA 240  
 DB 181 GTATCCGTACCGACGATGACGCGGAGAGAGCTTCAAACTCATCAACAGCAGTTAA 240  
 QY 241 GTGCGCTGCTGCTGACATCACTTCACTATGCGATTCGCTGAAAGTACGGAATAC 300  
 DB 241 GTGCGCTGCTGCTGACATCACTTCACTATGCGATTCGCTGAAAGTACGGAATAC 300  
 QY 301 GGGTGGATTCCTGCTGTTAACTCGGCAATATGGAATGAAGAGGTATTCGATG 360  
 DB 301 GGGTGGATTCCTGCTGTTAACTCGGCAATATGGAATGAAGAGGTATTCGATG 360  
 QY 361 GTGATTGACGTGCGGCGCATAAATTCGATCCGATTGGCGTTAAGCGCGGATG 420  
 DB 361 GTGATTGACGTGCGGCGCATAAATTCGATCCGATTGGCGTTAAGCGCGGATG 420  
 QY 421 CTGAAAAAGATCTGCAAGAAAAGATGCGGACCGACCGCGCGGCTGCGAATCT 480  
 DB 421 CTGAAAAAGATCTGCAAGAAAAGATGCGGACCGACCGCGCGGCTGCGAATCT 480  
 QY 481 GCCATGGCTCATGTTGATCTCTGATCGCTGAATTTCCATCTCAAGTCAAGCTG 540  
 DB 481 GCCATGGCTCATGTTGATCTCTGATCGCTGAATTTCCATCTCAAGTCAAGCTG 540  
 QY 541 AAGCGCTGACGCTCTCTCGCTGTTGAGCTTATGTTTGTGCGCAAAACAGATCAT 600  
 DB 541 AAGCGCTGACGCTCTCTCGCTGTTGAGCTTATGTTTGTGCGCAAAACAGATCAT 600  
 QY 601 CAGCGCTGATCTGGGATCAACGAAAGCGGTGTCGCGCGACGGGAGTAAATCC 660  
 DB 601 CAGCGCTGATCTGGGATCAACGAAAGCGGTGTCGCGCGACGGGAGTAAATCC 660  
 QY 661 GCCATTGGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGCGCTATCGCTG 720  
 DB 661 GCCATTGGTTAGTCTGCTGCTGTCTGAAGGATCGGCGACACGCTGCGCTATCGCTG 720  
 QY 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780  
 DB 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780  
 QY 781 CGTTCCGAGGATCAACTTCATCGCTCCGACCTGTCGCTCAGAAATTTGATGTT 840  
 DB 781 CGTTCCGAGGATCAACTTCATCGCTCCGACCTGTCGCTCAGAAATTTGATGTT 840

GPB AG GENOME PHARMACEUTICALS (DE)  
Location/Qualifiers  
1. 1119  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
BASE COUNT 272 a 279 c 317 g 251 t  
ORIGIN  
Query Match 100.0%; Score 1119; DB 6; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 8.6e-266;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCATACCAAGGCTCCATTCACGTAAGAAATCAACAGGATTTACGTTGGGAATGTG 60  
DB 1 ATGCATACCAAGGCTCCATTCACGTAAGAAATCAACAGGATTTACGTTGGGAATGTG 60  
QY 61 CCGATTGGCGATGTGTCCTCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120  
DB 61 CCGATTGGCGATGTGTCCTCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120  
QY 121 GTCCAGCAACCGGTCAATCAAAATCAAGCGCTGGAGCGGTGGCGCTGATATGTCCTG 180  
DB 121 GTCCAGCAACCGGTCAATCAAAATCAAGCGCTGGAGCGGTGGCGCTGATATGTCCTG 180  
QY 181 GTATCCGATCCGACGATGAGCGCGCGCAAGCGTTCAACTCATCAACAGCAGTTAAC 240  
DB 181 GTATCCGATCCGACGATGAGCGCGCGCAAGCGTTCAACTCATCAACAGCAGTTAAC 240  
QY 241 GTGCGCGTGGGTGATGATCATCTGATGATGATGATGATGATGATGATGATGATGATG 300  
DB 241 GTGCGCGTGGGTGATGATCATCTGATGATGATGATGATGATGATGATGATGATGATG 300  
QY 301 GCGCTCATTTCTGCGGTATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 360  
DB 301 GCGCTCATTTCTGCGGTATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 360  
QY 361 GTGCTGATCTGTGCGCGGATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 420  
DB 361 GTGCTGATCTGTGCGCGGATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 420  
QY 421 CTGGAAGAAAGATCTGCAAGAAAGATGAGCGAGCGCGGCGGTGCTGGAATCT 480  
DB 421 CTGGAAGAAAGATCTGCAAGAAAGATGAGCGAGCGCGGCGGTGCTGGAATCT 480  
QY 481 GCCATGCGTCAATGATGATCTGATGCGCTGGAATTCGATCAATCAAGTCAAGCTG 540  
DB 481 GCCATGCGTCAATGATGATCTGATGCGCTGGAATTCGATCAATCAAGTCAAGCTG 540  
QY 541 AAAGCGTCTGACGCTTCTCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 600  
DB 541 AAAGCGTCTGACGCTTCTCTGCTGCTGATGATGATGATGATGATGATGATGATGATG 600  
QY 601 CAGCGTTCATCTGCGGATCAAGAGCGGTGCGCGAGCGGCGGCGGAGTAAATTC 660  
DB 601 CAGCGTTCATCTGCGGATCAAGAGCGGTGCGCGAGCGGCGGCGGAGTAAATTC 660  
QY 661 GCCATGCTGATGATGCTGCTGCTGCTGATGAGCATTCGCGCAACGCTGCGTATCCTG 720  
DB 661 GCCATGCTGATGATGCTGCTGCTGCTGATGAGCATTCGCGCAACGCTGCGTATCCTG 720  
QY 721 GCGCGCGATCTGCTGCTGCTGCTGCTGATGAGCATTCGCGCAACGCTGCGTATCCTG 780  
DB 721 GCGCGCGATCTGCTGCTGCTGCTGCTGATGAGCATTCGCGCAACGCTGCGTATCCTG 780  
QY 781 CGTTGCGAGGAGATCAATCTGATGCGCTGCGCGCATGCTGCGGATTTGATGTT 840  
DB 781 CGTTGCGAGGAGATCAATCTGATGCGCTGCGCGCATGCTGCGGATTTGATGTT 840  
QY 841 ATCCGATCGTTAAACGCGCTGAGCAACGCTGGAAGATATCATCACTCCGATGAGCTT 900  
DB 841 ATCCGATCGTTAAACGCGCTGAGCAACGCTGGAAGATATCATCACTCCGATGAGCTT 900

QY 901 TCGATTATCGGCTGCGTGTGATGAGCCGAGGTGAGCGCGTGTTCATCACTCGGCGTC 960  
DB 901 TCGATTATCGGCTGCGTGTGATGAGCCGAGGTGAGCGCGTGTTCATCACTCGGCGTC 960  
QY 961 ACCGCGGCAACAGAAAGAGCGGCTCTATGAGATGCGGTGGCAAGACCGCTTGGAC 1020  
DB 961 ACCGCGGCAACAGAAAGAGCGGCTCTATGAGATGCGGTGGCAAGACCGCTTGGAC 1020  
QY 1021 AACACGATATGATGACACGCTGAGAGCAGCATTCGTCGCAAGCCGATGAGTAC 1080  
DB 1021 AACACGATATGATGACACGCTGAGAGCAGCATTCGTCGCAAGCCGATGAGTAC 1080  
QY 1081 GAAGCGGTGCAATGACGTTGAGTGAATAA 1119  
DB 1081 GAAGCGGTGCAATGACGTTGAGTGAATAA 1119  
RESULT 2  
AX038912 1119 bp DNA linear PAT 16-NOV-2000  
LOCUS  
DEFINITION Sequence 29 from Patent W00061793.  
ACCESSION AX038912  
VERSION AX038912.1 GI:11228221  
KEYWORDS  
SOURCE  
ORGANISM  
Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
1  
AUTHORS Lofler, H. and Jacob, A.  
TITLE Novel method for identifying antibacterial compounds  
JOURNAL Patent: WO 0061793-A 29 19-OCT-2000;  
GPB BIOTECH AG (DE); LOFLER HANNES (DE); JACOBI ALEXANDER (DE)  
FEATURES  
source  
1. 1119  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
BASE COUNT 272 a 279 c 317 g 251 t  
ORIGIN  
Query Match 100.0%; Score 1119; DB 6; Length 1119;  
Best Local Similarity 100.0%; Pred. No. 8.6e-266;  
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGCATACCAAGGCTCCATTCACGTAAGAAATCAACAGGATTTACGTTGGGAATGTG 60  
DB 1 ATGCATACCAAGGCTCCATTCACGTAAGAAATCAACAGGATTTACGTTGGGAATGTG 60  
QY 61 CCGATTGGCGATGTGTCCTCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120  
DB 61 CCGATTGGCGATGTGTCCTCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120  
QY 121 GTCCAGCAACCGGTCAATCAAAATCAAGCGCTGGAGCGGTGGCGCTGATATGTCCTG 180  
DB 121 GTCCAGCAACCGGTCAATCAAAATCAAGCGCTGGAGCGGTGGCGCTGATATGTCCTG 180  
QY 181 GTATCCGATCCGACGATGAGCGCGCGCAAGCGTTCAACTCATCAACAGCAGTTAAC 240  
DB 181 GTATCCGATCCGACGATGAGCGCGCGCAAGCGTTCAACTCATCAACAGCAGTTAAC 240  
QY 241 GTGCGCGTGGGTGATGATCATCTGATGATGATGATGATGATGATGATGATGATGATG 300  
DB 241 GTGCGCGTGGGTGATGATCATCTGATGATGATGATGATGATGATGATGATGATGATG 300  
QY 301 GCGCTCATTTCTGCGGTATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 360  
DB 301 GCGCTCATTTCTGCGGTATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 360  
QY 361 GTGCTGATCTGTGCGCGGATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 420  
DB 361 GTGCTGATCTGTGCGCGGATTAACCTGCGCAATTCGATTAAGAGCGATTCGATG 420

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 13:04:06 ; Search time 4350 Seconds

(without alignments)  
10523.649 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119  
Sequence: 1 atgcataaccagagctccat.....ttcagcaggttgaaaataa 1119Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapept 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_man:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_other:*
41: em_hcg_hum:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	6	AX036302 Sequence
2	1119	100.0	1119	6	AX038912 Sequence
3	1119	100.0	1119	6	AX050487 Sequence
4	1119	100.0	1119	6	AX039393 Sequence
5	1119	100.0	1697	1	ECGCEP
6	1119	100.0	13176	1	AE000338 Escherich
7	1098.4	98.2	10225	1	AE015271 Shigella
8	1098.4	98.2	290380	1	AE016987 Shigella
9	1098.2	98.1	300099	1	AE016764 Escherich
10	1093.4	97.7	1119	1	AY033515 Escherich
11	1093.4	97.7	11521	1	AE005481 Escherich
12	1093.4	97.7	296827	1	AP002561 Escherich
13	1010	90.3	1010	6	AR027993 Sequence
14	1010	90.3	1010	6	AR028328 Sequence
15	1010	90.3	1010	6	BD084908 Methods f
16	887	79.3	23647	1	AE008814 Salmonell
17	885.4	79.1	145050	1	AL627275 Salmonell
18	885.4	79.1	300431	1	AE016835 Salmonell
19	885.4	79.1	300431	1	AE016835 Salmonell
20	690	61.7	16214	1	D90880 E.coli geno
21	685.4	61.3	65821	2	AC020826 Nus muscu
22	665	59.4	11824	1	AE004161 Vibrio ch
23	658.4	58.8	301235	1	AP005075 Vibrio pa
24	632.4	56.5	301442	1	AE016798 Vibrio pa
25	631.6	56.4	1345	1	PSU67933 Providencia
26	631.6	56.4	1345	6	AR027991 Sequence
27	631.6	56.4	1345	6	AR028326 Sequence
28	631.6	56.4	1345	6	BD084906 Methods f
29	622	55.6	10507	1	U32721 Haemophilus
30	622	55.6	110000	6	AR274513_03 Continuation (4 of
31	620.6	55.5	11161	1	AE013738 Yersinia
32	620.6	55.5	204050	1	AJ414154 Yersinia
33	619.8	55.4	1079	6	AR027992 Sequence
34	619.8	55.4	1079	6	AR028327 Sequence
35	619.8	55.4	1079	6	BD084907 Methods f
36	603.6	53.9	11268	1	AE006237 Pasteurel
37	566	50.6	10607	1	AE015769 Shewanell
38	534.6	47.8	886	6	AR027995 Sequence
39	534.6	47.8	886	6	AR028330 Sequence
40	534.6	47.8	886	6	BD084910 Methods f
41	533	46.9	300861	1	AE016777 Pseudomon
42	525	46.9	311249	1	AE016860 Pseudomon
43	515.4	46.1	17148	1	AE004798 Pseudomon
44	403.6	36.1	110000	6	BD061520_3 Continuation (4 of
45	403.6	36.1	347550	1	AP001118 Buchnera

## ALIGNMENTS

```
RESULT 1
LOCUS AX036302 1119 bp DNA
DEFINITION Sequence 29 from Patent EP1043403.
ACCESSION AX036302
VERSION AX036302.1 GI:11225912
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Novel method for identifying antibacterial compounds
PATENT: EP 1043403-A 29 11-OCT-2000;
```





QY 307 ArglyGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGly 326  
 125 CGAAGAGGTGGCTGGCCGCCCACTCATCAAGCTGTAATTTTACAGATTAATGGGA 184  
 QY 327 PheIleIleGlnGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGly 346  
 185 TTTATCATTTGACGAGATATATTAACCGAATTTGCCGAGCTCAGACATCCATTTTGAGGA 244  
 QY 347 AspLeuLeuTyrAlaIleGlnArgValLeuLysLeuSerValProAsnLeuTyrValTyr 366  
 245 AACCTTTTACGCCCATCCAGAGGTAATGAACTTTCAGTTCCATTTATATGCTGG 304  
 QY 367 LeuCysMetPheTyrCysPhePheHisLeuTyrLeuAsnIleLeuAlaGluLeuLysCys 386  
 305 CTTGTCATGTTCTACAGCTCTTTCATCTTTGTTAAATATCTTCGGAAGTTCTGGCA 364  
 QY 387 PheGlyAspArgGlnPheTyrLysAspTyrPrrPaspAlaLysSerValGlyAspTyrTyr 406  
 365 TTTGGGATCGTGAAGTCTATTAAGAATGTGTGACGCAAAACAAATGATGATGCTG 424  
 QY 407 ArgMetPrrPaspMetProValHisLysTrrMetValArgHisIleTyrPheProCysLeu 426  
 425 AGACTTTGGAATATGCTGCTGATAGTGAATGATGCTGCTGACATCTATTTCCATGCTTA 484  
 QY 427 ArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAla 444  
 485 AGAAATGACATACCTAAGGAGGAGTGCATAGTGCATCTCTTTTATATCTGCT 538

## RESULT 12

Bj316561

LOCUS Bj316561 606 bp mRNA linear EST 09-APR-2002  
 DEFINITION BstJ16561 Y. Ogihara unpublished cDNA library, Wh\_yf Trilicium  
 accession cDNA clone whyf23117 5', mRNA sequence.

Bj316561

Bj316561.1 GI:20121748

EST

Trilicium aestivum (bread wheat)

Trilicium aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

Triliceae; Trilicium.

1 (bases 1 to 606)

Ogihara, Y. and Mural, K.

Unpublished

Expressed genes in Trilicium aestivum

Contact: Tadasu Shin-1

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@genes.nig.ac.jp.

Location/Qualifiers

1: 606

/organism="Trilicium aestivum"

/mol\_type="mRNA"

/cultivar="Chinese Spring"

/db\_xref="taxon:4565"

/clone="whyf23117"

/tissue\_type="spikelet at early flowering"

/dev\_stage="Peekes" scale 6"

/clone\_lib="Y. Ogihara unpublished cDNA library, Wh\_yf"

BASE COUNT 170 a 113 c 125 g 198 t

ORIGIN

Alignment Scores:

Pried. No.: 1.22e-63

Percent Similarity: 762.50

Best Local Similarity: 79.21%

Query Match: 67.33%

DB: 27.52%

US-09-623-514a-2 (1-520) x Bj316561 (1-606)  
 QY 238 GlyValThrLeuMetLeuThrCysIleValTrrLeuLysLeuValSerTyrAlaHis 257  
 1 GGTTTTGTGATGATGTTTATGCTGCTCATGTTGGCTGAGCTCATCTTTTGGCCAT 60  
 QY 258 ThrSerTyrAspIleArgSerLeuAla-----AsnAlaAlaAsp 270  
 61 ACAACCATGACATTAAGCAATTAACCATTAAGTATGATATGACACCAC 120  
 QY 271 LysAlaAsnProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetVal 290  
 121 ACAGCTGACATGATGATTAATTAACAGCTCCAACTTTGAGGACCTTATATACCTCAAGANG 180  
 QY 291 AlaProThrLeuLysTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTyr 310  
 181 GCTCCACACCTGGTGCATACGCCAGCTTAATCTCGAAGTGAATAATGTGAAAAAGGCTGG 240  
 QY 311 ValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIleGlu 330  
 241 CTGATTCGCAATATTAATTCGCACCTGATATTAAGGATTAACAGTTTCATTAATGAG 300  
 QY 331 GlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuTyr 350  
 301 CAGTACATTAATCCAACTGCTGGAACCTCCAACTCCATTCGAAAGCGGACTTCGAAAT 360  
 QY 351 AlaIleGlnArgValLeuLysLeuSerValProAsnLeuTyrValTrrLeuCysMetPhe 370  
 361 GCTGTGAGACAGCTTTGTGAGGCTTCATACCAAAAGTTTACTTAAGCTTTCACAGTTC 420  
 QY 371 TyrCysPhePheHisLeuTrrPleuAsnIleLeuAlaGluLeuLysPheGlyAspArg 390  
 421 TATTCCTTTTTCATCTGCTGTTAAATATTCCTGAGATCTTCGTTGCTGGTGACCG 480  
 QY 391 GluPheTyrLysAspTrrPrrPaspAlaLysSerValGlyAspTrrPargMetPrrPasn 410  
 481 GAATTCACAAACACGTTGGAATGCAAAACAAATGAGATGATGCGGCAAAATGGAAC 540  
 QY 411 MetProValHisLysTrrMetValArgHisIleTyrPheProCysLeuSerLysIle 430  
 541 ATGCCCGTCATTAATGATGTTGCTGCTCATATATATTTCTTGCATGCGAAGTGGTATA 600  
 QY 431 ProLys 432  
 601 TCGAAG 606

## RESULT 13

AM349274/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

EST

glycine max (soybean)

glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids

; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

glycine.

1 (bases 1 to 855)

Vodkin, L., Keln, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,

Expanding, D., Raph, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.

A Functional Genomics Program for Soybean (NSF 9872565)

Unpublished

Other-ESTs: A1441040

Contact: Vodkin, L.O., PI, A Functional Genomics Program for

Soybean (NSF 9872565)

Lewin, H. A., Director, Keck Center for Comparative and Functional

Genomics

University of Illinois

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA

```

VERSION      BJ322102.1  GI:20125036
KEYWORDS     EST
SOURCE       Triticum aestivum (bread wheat)
ORGANISM     Triticum aestivum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
              ; Triticeae; Triticum
              1 (bases 1 to 729)
REFERENCE    Ogihara, Y. and Murai, K.
              Expressed genes in Triticum aestivum
              Unpublished
JOURNAL      Contact: Tadasi Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshin@genes.nig.ac.jp.
FEATURES     Location/Qualifiers
              1..729
              /organism="Triticum aestivum"
              /mol_type="mRNA"
              /cultivar="Chinese Spring"
              /db_xref="taxon:4565"
              /clone="wbf23117"
              /tissue_type="Spikelet at early flowering"
              /dev_stage="Peekes" scale 6"
              /clone_1lb="Y. Ogihara unpublished cDNA library, wh_lyf"
BASE COUNT   237 a 146 c 153 g 193 t
ORIGIN
Alignment Scores:
Pred. No.:    2,53e-70      Length:      729
Score:        832.50       Matches:    147
Percent Similarity: 84.13%   Conservative: 28
Best Local Similarity: 70.67% Mismatches: 30
Query Match:  30.04%       Indels:     3
DB:           12          Gaps:       1
US-09-623-514A-2 (1-520) x BJ322102 (1-729)
OY 313 ArgGlnPheAlaIbysLeuValIlePheThrGlyPheMetGlyPheIleIleGlu-GLNty 332
DB 729 GGTCAAAATTATTCGCACTGTATATTACAGGTATTCAGAGTTTCATTATTCAGAACAGTA 670
OY 332 rIleasnProIleValaIcAasnSer-LysHisProLeuLysGlyAspLeuLeuTyraIar 352
DB 669 CATAAATCCANTCGTCGAGACTCTCCACATCCATTCGAAGCGGAGCTTCGAATGCTG 610
OY 352 legIuArgValleuLysLeuSerValProasnLeuTyraIarIlePheLysCysMetPheTyrc 372
DB 609 TAGAGACTGTCTCGAGCTCATACCAAAATGTTACTTATAGGCTTGACATGTTCTATAT 550
OY 372 yspPheHisLeuTrpPheLeuasnIleleuAlaGluLeuLysCysheGlyAspArgGluP 392
DB 549 GCTTTTCCATCTCGTGGTTAAATATCTTCGAGATTCCTCGTTGCTGCGTACCGTGAAT 490
OY 392 heTyryLysAspTrpTrpAsnAlaLysSerValGlyAspTyrrTrpArgMetTrpAsnMetP 412
DB 489 TCTACAAAGACTGGGGAATGCAAAACAAATTGATGAGTATTTGGGAAATGCAACATCGC 430
OY 412 rovalHisLysTrpMetValaIcArgHisIleTyrrPheProCysLeuArgSerLysIleProL 432
DB 429 CCGTCGCAAAATGATGTCTCGTCATATATATTTTCCTTCGATCGGAGTGGTATATCGA 370
OY 432 ysrHleuAlaIleIleIleAlaPheLeuValSerAlaValaPheHisGluLeuCysIleA 452
DB 369 AGGAAGTGTCTGTTTGTATCATTTTGTATTCGCCGTGCTCCATAGACTTGTGTG 310
OY 452 laValProCysArgLeuPheLysLeuTyrrAlaPheLysGlyrIleMetheGlnValProL 472
DB 309 CTGTCCCTCGCGGAATTCGTGAAGTTCGTGGGCAATTCCTAGGAGTCAAGTGTGAGATCCCC 250

```

```

OY 472 euValaPheIleThrAsnTyrrLeuGlnGluAqPheGlySerTherVal---GlyasnMetI 491
DB 249 TTATCAATTCACATCATGCGTCAAAACCAAAATTCAGGATACATGATGCCCGCAACATGA 190
OY 491 lePheTrpPheIlePheCysIlePheGlyGlnProMetCysValleuLeuTyrrHisA 511
DB 169 TATTTGGTCTTTTTCGATCTAGCGACACCTATGCTGTCTCTCATACCATG 130
OY 511 spleuMetAsnArgLysGly 517
DB 129 ATGTGATGAACAGCATGTGGC 110
RESULT 11
LOCUS       BI422212                      539 bp    mRNA    linear    EST 16-AUG-2001
DEFINITION  EST532878 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            c16c69M10 5' end, mRNA sequence.
ACCESSION   BI422212
VERSION     BI422212.1  GI:15196417
KEYWORDS    EST.
SOURCE      Lycopersicon esculentum (tomato)
ORGANISM   Lycopersicon esculentum
REFERENCE   Alcala, J., Vrebalov, J., White, R., Matern, A. L., Vision, T., Holt, I. E.,
            Jiang, F., Upton, J., Craven, M. B., Bowman, C. L., Ahn, S., Ronning,
            C. M., Fraser, C. M., Martin, G. B., Tanksley, S. D. and Giovannoni, J.
            Generation of ESTs from tomato callus tissue
            Unpublished
JOURNAL     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html.
FEATURES     Location/Qualifiers
              1..539
              /organism="Lycopersicon esculentum"
              /mol_type="mRNA"
              /cultivar="TA96"
              /db_xref="taxon:4081"
              /clone="c16c69M10"
              /tissue_type="callus"
              /dev_stage="25-40 days old"
              /lab_host="XLI-Blue MRF"
              /clone_1lb="tomato callus, TAMU"
              /note="Vector: pBluescript SK(-); Site 1: EcoRI, Site 2:
              XhoI; supplier: Giovannoni laboratory; c16c - Cotyledons
              of seedlings 7-10 days post-germination were excised, cut
              at both ends and placed on MS medium with no selection.
              Mixed callus was harvested at 25 and 40 days and included
              undifferentiated masses. Tomato Callus EST library"
BASE COUNT   137 a 102 c 120 g 180 t
ORIGIN
Alignment Scores:
Pred. No.:    5.91e-64      Length:      539
Score:        765.00       Matches:    136
Percent Similarity: 87.08%   Conservative: 19
Best Local Similarity: 76.40% Mismatches: 23
Query Match:  27.61%       Indels:     0
DB:           12          Gaps:       0
US-09-623-514A-2 (1-520) x BI422212 (1-539)
OY 267 AsnaIaAAspLyAsIaAsnProGluValSerTyrrTyrrValSerLeuLysSerLeuA 286
DB 5 AATGAGGTGAGAAATTCGGAATCAATCACTTACATAGTTTGGAGAGTTGGCT 64
OY 287 TyrPheMetValaIaProThrLeuCysTyrrGlnProSerTyrrProArgSerAlaCysIle 306
DB 65 TACTTCAGAGTTGCTCCAACTTATGCTATACGCTTACCTACCTCGCTCCATCCAT 124

```







REFERENCE  
1 (bases 1 to 813)  
AUTHORS  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
TITLE  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
JOURNAL  
Unpublished  
COMMENT  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aac cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 16 row: H column: 13  
Seq primer: gta aac cga cgg cca gtc.  
Location/Qualifiers  
1. 813

FEATURES  
source  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNE0219"  
/tissue\_type="leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNEB"  
/note="Vector: pBluescript II KS +; Site\_1: EcoRI; site\_2:  
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061  
ORIGIN  
BASE COUNT 211 a 147 c 165 g 290 t

## Alignment Scores:

Pred. No.: 1.39e-82 Length: 813  
Score: 959.50 Matches: 169  
Percent Similarity: 76.95% Conservative: 38  
Best Local Similarity: 62.83% Mismatches: 55  
Query Match: 34.63% Indels: 7  
DB: 14 Gaps: 1

US-09-623-514a-2 (1-520) x CB651490 (1-813)

OY 156 MetLysTyrGlyTrpLeuIleArgThrAspPheSerSerArgSerLeuArgAsp 175  
DB 6 TTGCAGATAGGCTTATTATTAAGAGCTGGGTTTGGTTTATGATTAATCATTCATGGCGGAC 65  
OY 176 TrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaIlePheThrVal 195  
DB 66 TGGCCACTTCTATATGTTGTCTTAATGCTCCGCTTCCCTCCGCTGCATTTGCAGTT 125  
OY 196 GluLysLeuValLeuGlnLysTyrTrpIleSerGluProValIlePheLeuHsIleIle 215  
DB 126 GAAGAAGTTGGCATTAACATGTTATCTAGATGCTGCTACTCCCTCCATATCTTC 185  
OY 216 IleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPhe 235  
DB 186 CTTTCAACAACGCAAAATGATATACAGTCGCTGTGATTTCTTAAAGTGTGATTCGAGTT 245  
OY 236 LeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyr 255  
DB 246 TTGCTGGGCTTTTGTGATATTTATTTGCTGATATGCTTGGCTGAAGCTTGATCTTTT 305  
OY 256 AlaHisThrSerTyrTrpSerIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGlu 275  
DB 306 GCACATACAAACCATGATATTAAGCAACGACCATGGCGGCAAGAGTTGATTAATGA 365  
OY 276 ValSerTyrTyr-----ValSerLeuLysSerLeuAlaTyrPhe 288  
:::||||| ::||| ::||| ::||| ::|||

DB 366 CTAAAGCAGCTTGACATGATTAATTTACAACTCCCACTTAGGAATCTAATACTTC 425  
OY 289 MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLys 308  
DB 426 ATGATGCTCTTACCTCTTATATCACCACCAAGCTATCCCGAAGCTTATGTTAGAAA 485  
OY 309 GlyTrpAlaAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle 328  
DB 486 GGTGGCTGATGTGCATAATATATCTGACTGATGATCTTTACTGCTTCAAGGCTTCAT 545  
OY 329 IleGlnIleThrIleAsnProIleValAlaArgAsnSerLysHisProLeuLysGlyAspLeu 348  
DB 546 ATGAGCAATACATTAATTCATATGTTGATTCAGATTCATCATGAAAGGAGGACTC 605  
OY 349 LeuTrpAlaIleGluAlaGValLeuLysLeuSerValProAsnLeuTyrValTrpLeuCys 368  
DB 606 CTAAATGCTGTAGAGACTGTTTGAACCTCTCATTTACCAATGTTTACCTGTGGCTTGC 665  
OY 369 MetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCysPheGly 388  
DB 666 ATGTTCTATGCTTTTTCATCTCTGTTAGTAACTGCTGATGATTCGATTTGGT 725  
OY 389 AspArgGluPheThrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMet 408  
DB 726 GACCGGAAATCTTACAAAGATGGTGGAATGCACAAATGATGATGAGATAAGAGAAA 785  
OY 409 TrpAsnMetProValHisLysTrpMet 417  
DB 786 TGGATATAGCTGTGATCAATAATGGTT 812

RESULT 7  
CB660636 851 bp mRNA linear EST 09-APR-2003  
LOCUS OSJNE0219.f OSJNE0219 5', mRNA sequence.  
DEFINITION clone OSJNE0219 5', mRNA sequence.  
ACCESSION CB660636  
VERSION CB660636.1 GI:29664361  
KEYWORDS EST.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Enarthroidae; Oryzaceae; Oryza.  
1 (bases 1 to 851)  
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,  
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.  
TITLE  
Large-scale identification of ESTs involved in the interaction  
between rice and Magnaporthe grisea  
JOURNAL  
Unpublished  
COMMENT  
Contact: Rod Wing  
Arizona Genomics Institute  
University of Arizona  
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ  
85721-0088, USA  
Tel: 520 626 3967  
Fax: 520 621 9288  
Email: http://genome.arizona.edu  
PCR Primers  
FORWARD: gta aac cga cgg cca gtc  
BACKWARD: gga aac agc tat gac cat g  
Plate: 02 row: J column: 19  
Seq primer: gta aac cga cgg cca gtc.  
Location/Qualifiers  
1. 851

## FEATURES

source  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="mRNA"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="OSJNE0219"  
/tissue\_type="leaf"  
/dev\_stage="3 week"  
/lab\_host="DH10B"  
/clone\_lib="OSJNEB"

```

OY 387 phegiyaparggluphetyrlyasprtrpasaalalyserValglysptyrtrp 406
    |||||||
DB 385 TTGGGGAGTCGAGCTTCTATTAAGATTGGTGGAAAGCAATTCATGATGATGG 424
OY 407 ArgmetrPasnmeProvalHisIlystrpMetValArgHisIleTyPheProCysLeu 426
    |||||||
DB 425 AGACTTGGANATGCCCTGTACATAGTGAGTTCGTCACATCATATTTCCACGCTTA 484
OY 427 ArgSerIyIleProIystrpLeuValIleIleIleIleValPheLeuValSerIalValPhe 446
    |||||||
DB 485 AGAATGCATACATCAAGGAGTGCATGCTCTTCTTATATCATCTGCTGTTTC 544
OY 447 HIsGluLeuCysIleIalValProCysArgLeuPheIystrpLeuValPheLeuGlyIle 466
    |||||||
DB 545 CATGAGCTGTGTATGCTGTCTGTGCGCTATTCAGTTTGGCATCTCTTGGAATC 604
OY 467 MetPheGlnValProLeuValPheIleThrAsnTyrlaGlnIArgPhe---GlySer 485
    |||||||
DB 605 ATGTTTCAGATTCCTTGTGTCTACTACAGAACTTCCTGCACAAAGTTCAAAATCTCG 664
OY 486 ThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnPrometCysVal 505
    |||||||
DB 665 AATGGGCAACATGACATCTGCTGCTTTCTCTGATGTCATCAACCATG-TGTGTG 723
OY 506 LeuLeu 507
    ||||||
DB 724 CTCTGTG 729

```

```

RESULT 5
CB660599 872 bp mRNA linear EST 09-APR-2003
LOCUS OSJNEB02118.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEB02118 5', mRNA sequence.
ACCESSION CB660599
VERSION CB660599.1 GI:29664324
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euphorbiales; Oryzae; Oryza.
1 (bases 1 to 872)
JantasaIyaret,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cga cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 02 row: 1 column: 18
Seq primer: gta aac cga cga cca gtc.
Location/Qualifiers
1. 872

```

```

FEATURES
Source
1. 872

```

```

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB02118"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_id="OSJNEB"
/note="Vector: pBluescript II KS+. Site.1: EcoRI, Site.2:
XhoI; 24 hrs after inoculation with Rice Blast (CG240-1)"

```

```

BASE COUNT 227 a 174 c 172 g 299 t
ORIGIN
Alignment Scores:
Pred. No.: 3,2e-83 Length: 872
Score: 966.50 Matches: 162
Percent Similarity: 87.05% Conservative: 33
Best Local Similarity: 72.32% Mismatches: 28
Query Match: 34.88% Indels: 1
DB: 14 Gaps: 1
US-09-623-514a-2 (1-520) x CB660599 (1-872)

```

```

OY 293 ThrLeuCysTyrlaGlnProSerTyrlaProArgSerIalCysIleArgIySGlyTrpValAla 312
    |||||||
DB 6 ACACCTGTTATACAGCAAGTATCCCGACACTCATGTTGAGAAAGTTGGCTGAT 65
OY 313 ArgGlnPheAlaIystrpValIlePheThrIyGlyPheMetGlyPheIleIleGlyIle 332
    |||||||
DB 66 CGTCAAAATTAATCTGTACTGATCTTACTGCTGCTTCAAGGCTTCATTTAGCAATAC 125
OY 333 IleAsnProIleValArgAsnSerIySHsProLeuIyGlyAspLeuLeuTyrlaIle 352
    |||||||
DB 126 ATAAATCCATGTTGTGTGATTCACGACATCCATTCATTAAGAGAGACTCTTAATGCTGTA 185
OY 353 GluArgValLeuIystrpLeuSerValProAsnLeuTyrlaTyrlaPheCysMetPheTyrcys 372
    |||||||
DB 186 GAGACTGTTTGAACCTCATACCAAAATGTTACCTGTGCTTGCATGTTCTATGCT 245
OY 373 PhePheHisLeuTrpLeuAsnIleLeuAlaGlyLeuLeuCysPheGlyAspArgGlyPhe 392
    |||||||
DB 246 TTTTTCACATCTGTGTGATTAATCTGCTGCTGATTCGATTCGATTCGATTCGATTCGATTC 305
OY 393 TyrlaAspTrpTrpPasnAlaIystrpValIyAspTyrlaArgMetPasnMetPro 412
    |||||||
DB 306 TCAAAAGATTTGGTGGATCCAAACAAATTCATGATTCGATTCGATTCGATTCGATTCGATTC 365
OY 413 ValHisIystrpMetValArgHisIleTyrlaPheProCysLeuArgSerIyIleProIy 432
    |||||||
DB 366 GTACATAAATGGGTGTGTCGATTAATTCCTTCTGTCATGCAGAAATGATATACAAAG 425
OY 433 ThrLeuAlaIleIleIleIleIleIleValPheLeuValSerIalValPheHisGlyLeuCysIleAla 452
    |||||||
DB 426 GAAGTGTGCTGTGATGATCAATTCCTGTTTGTGCGGATCTCCAGATGATGATGTCGCT 485
OY 453 ValProCysArgLeuPheIystrpLeuTrpPheIleGlyIleMetPheGlnValProLeu 472
    |||||||
DB 486 GTTCCCTGGCGCATCTCAAGTTCGCGCATCTTAAGAAATATCTACAGATCCCTTT 545
OY 473 ValPheIleThrAsnTyrlaGlnIArgPheGlySerThr---ValGlyAsnMetIle 491
    |||||||
DB 546 ATCGTATTCACAGCATACCTCAAAAGTAATTCAGAGATACATGATGGTGGCAACATGATA 605
OY 492 PheTrpPheIlePheCysIlePheGlyGlnPrometCysValLeuLeuTyrlaHisAsp 511
    |||||||
DB 606 TTTTGTCTTTTTCGATTCATGATGCGCAGCAAGTGTCTCTCTGATCATCATGAT 665
OY 512 LeuMetAsnArg 515
    |||||||
DB 666 GTGATGACACAGG 677

```

```

RESULT 6
CB651490 813 bp mRNA linear EST 08-APR-2003
LOCUS OSJNEB16H13.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
DEFINITION clone OSJNEB16H13 5', mRNA sequence.
ACCESSION CB651490
VERSION CB651490.1 GI:29646483
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

## ORIGIN

## Alignment Scores:

Pred. No.: 7,69e-93 Length: 1181  
 Score: 1067.50 Matches: 192  
 Percent Similarity: 79.71% Conservative: 28  
 Best Local Similarity: 69.57% Mismatches: 47  
 Query Match: 38.52% Indels: 9  
 DB: 11 Gaps: 3

US-09-623-514a-2 (1-520) x AY110660 (1-1181)

QY 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAspLys--- 271  
 DB 12 GTCTTATGACATACAAATATGATATGAAGGATATGTCACAAATACCTAGGAAGGCT 71  
 QY 272 -----AlaAsnProGluValSerTyrTyrValSerLeuLysSerLeu 285  
 DB 72 GCTGCATATGAAATATATGTCATCCAGAAATATGAAGATCCACCTTTAAAGTCTA 131  
 QY 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305  
 DB 132 GTGTACTTCAATGTTGGCCCAACACTTGTTCACAGCAACTATCTCTCAACACTATGT 191  
 QY 306 IleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325  
 DB 192 ATTAGAAAGGGTGGTGGTACCCAGCAACATCATTAAGTGGCTGNNNNNACAGGCTGATG 251  
 QY 326 GlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345  
 DB 252 GGGCTTCAATATGAGCAATATATMAACCAATGTGAAAGATTCACCAACTCATAA 311  
 QY 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365  
 DB 312 GCGAANNNNNGAATGCTATAGAAAGAGTCTTAAACTCTCAGTGGCCCACTTATATGTA 371  
 QY 366 TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeu 385  
 DB 372 TGGCTTGCATGTTCTATTCNNNNNNNCACTTATAGCTGAACATTTAGCTGAACCTCTC 431  
 QY 386 CysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAlaLysSerValGlyAspTyr 405  
 DB 432 TGTTCGTCACCGTGAATCTATMAAGACTGGGGAATCCCAAACTGTTGAAGACTAC 491  
 QY 406 TrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisLysLysPheProCys 425  
 DB 492 TGGAGGATGTGGAACATGCTGTTCAATGATGATCATGACACATATATTTCCATGT 551  
 QY 426 LeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaVal 445  
 DB 552 ATTAGCAAGAGCTTTTCCAGGGGTAGCTATCTATCTATCTGTTCTGTTCCAGCTGTA 611  
 QY 446 PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly 465  
 DB 612 TTCCATGATGATATGATATGCGGTGCGTGCACATTTCAAAATCTGCGCATTTTCGGG 671  
 QY 466 IleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySer 485  
 DB 672 ATCATGTTTCAGATACCGTTGATCTTGACAAAGATATCTCCATGCTAGCTTCAACAT 731  
 QY 486 Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCys 504  
 DB 732 GTAATGTGGGCAACATGATATTTTGGTTC---TTCCGTTATAGTCGACAGCCGATGTGT 788  
 QY 505 ValLeuLeuTyrTyrHisAspLeuMetAsnArgLysLysSerMetSer 520  
 DB 789 GTCTTCTATATACATGAGGATCATGATACAGGAGCCAGGCAACT 836  
 RESULT 4  
 BI422326 729 bp mRNA linear EST 16-AUG-2001  
 LOCUS EST5523992 tomato callus, TANK Lycopersicon esculentum cDNA clone  
 DEFINITION GLECC69N10 5' end, mRNA sequence.

ACCESSION BI422326  
 VERSION BI422326.1 GI:15196624  
 KEYWORDS EST  
 SOURCE Lycopersicon esculentum (tomato)  
 ORGANISM Lycopersicon esculentum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.  
 1 (bases 1 to 729)  
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Visions,T., Holt,I.E.,  
 Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Roming,  
 C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D., and Giovannoni,J.  
 Generation of ESTs from tomato callus tissue  
 Unpublished  
 CONTACT: CUGI  
 Institution: Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Email: <http://www.genome.clemson.edu/orders/index.html>  
 Location/Qualifiers  
 1..729  
 /organism="Lycopersicon esculentum"  
 /mol\_type="mRNA"  
 /cultivar="TA496"  
 /db\_xref="taxon:4081"  
 /clone="GLECC69N10"  
 /tissue\_type="callus"  
 /dev\_stage="25-40 days old"  
 /lab\_host="Xil-Blue MRP"  
 /clone\_id="tomato callus, TANK"  
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:  
 XhoI; Supplier: GibcoBRL laboratory; GLEC - Cotyledons  
 of seedlings 7-10 days post-germination were excised, cut  
 at both ends and placed on MS medium with no selection.  
 Mixed callus was harvested at 25 and 40 days and included  
 undifferentiated masses. Tomato Callus EST library"

## FEATURES

## source

BASE COUNT 177 a 142 c 160 g 250 t  
 ORIGIN

## Alignment Scores:

Pred. NO.: 2.5e-88 Length: 729  
 Score: 1018.00 Matches: 185  
 Percent Similarity: 85.95% Conservative: 23  
 Best Local Similarity: 76.45% Mismatches: 33  
 Query Match: 36.74% Indels: 2  
 DB: 12 Gaps: 1

US-09-623-514a-2 (1-520) x BI422326 (1-729)

QY 267 AsnAlaIleAspLysAlaAsnProGluValSerTyrTyrValSerLeuLysSerLeuAla 286  
 DB 5 AATGAGGCTGAGATTCGCAATCACTACACTTATCTTTCAGAGTTGCT 64  
 QY 287 TyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIle 306  
 DB 65 TACTTCATGTGTGTCACACTTATGCTATACAGCTTATCTCTCTGCTCATTCATT 124  
 QY 307 ArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGly 326  
 DB 125 CGAAGGCTTGGCTGGCCCGCCCACTCATCAAGCTGTAATTTTACAGATTATAGGA 184  
 QY 327 PheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGly 346  
 DB 185 TTATCATTTGAGCATATATTAACCGATTTGCGAAGCTCACACATCCATTGAAGA 244  
 QY 347 AspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrp 366  
 DB 245 AACCTTTATACGCAATCAGAGAGGTATTGAAAGCTTTCAGTTCCAAATTTATATGCTGG 304  
 QY 367 LeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysCys 386  
 DB 305 CTCTGCATGTTCTACACCTCTTTCATCTTGTGTTAAATATCTGCGAAGTTCTCGCA 364

REFERENCE  
AUTHORS

1. (bases 1 to 961)  
Buell, C.R., Hart, A., Baker, B., Tanksley, S., Fry, W., Smart, C.,  
Restrepo, S., Griffiths, H., van der Hoeven, R., Tsai, J. and  
Karmycheva, S.A.

TITLE  
JOURNAL  
COMMENT  
Unpublished  
On Jun 10, 2002 this sequence version replaced gi:21369236.  
Other ESTs: EST617781

Contact: Robin Buell  
The Institute for Genomic Research  
9712 Medical Center Dr, Rockville, MD 20850, USA  
Email: potato-array@tigr.org  
This clone can be obtained from the University of Arizona Genomics  
Institute. Orders can be made through URL:  
http://genome.arizona.edu/orders/  
Seq primer: 17.

FEATURES  
source

Location/Qualifiers  
1..961  
/organism="Solanum tuberosum"  
/mol\_type="mRNA"  
/cultivar="Kennebec or Blonje"  
/db\_xref="taxon:4113"  
/clone="STMJ76"  
/tissue\_type="mixed tissues"  
/lab\_host="SOLR"  
/clone\_lib="Generation of a set of potato cDNA clones for  
microarray analyses mixed potato tissues"  
/note="Vector: pBluescript SK(+); Site 1: EcoRI; Site 2:  
XhoI; supplier: Combination of untreated and Phytophthora  
infestans-treated libraries of stolons, leaves, leaflets,  
axillary buds of stem explants, petioles, germinating eyes  
, tubers, or roots."

## BASE COUNT

324 a 205 c 181 g 251 t

## ORIGIN

324 a 205 c 181 g 251 t

## ALIGNMENT SCORES:

Pred. No.: 1 85e-101 Length: 961  
Score: 1155.00 Matches: 209  
Percent Similarity: 86.728 Conservative: 26  
Best Local Similarity: 77.128 Mismatches: 32  
Query Match: 41.688 Indels: 4  
DB: 13 Gaps: 3

US-09-623-514A-2 (1-520) x B0510367 (1-961)

254 SerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAspIleAla--- 272  
|||||  
961 TCTATGCAATCAATATGATATGATGACAGCTTGCAGAAAGCTGTGATGAGGTGAA 902  
273 AenProGluVal-----SerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetVal 290  
|||||  
901 AATTCGGAATCAACTACTCTTACATGTTAGTTGCAAGCTTGGCTTACTTCAATGATT 842  
291 AlaProThrLeuGlyTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTyr 310  
|||||  
841 GCTCCCACTTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATG 782  
311 ValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIleGlu 330  
|||||  
781 CTGGCCGGCAACTCAAGCTGATATTTTACAGATATATGAGATTATCATTTAGAG 722  
331 GlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLysTyr 350  
|||||  
721 CAGTATATTAACCCGATTTGCGAAGCTCACAACATCCATTGGAAGAAACCTTTATAC 662  
351 AlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTyrLeuLysMetPhe 370  
661 GGCATGGAAGGGATATGAAGCTTCAAGTTCAATTTATATGCTGCTGCTGCAATGTTTC 602  
371 TyrCysPhePheHisLeuTyrLeuAsnIleLeuAlaGluLeuLeuLysPheGlyAspArg 390  
601 TACAGCCTCTTTCATCTTTGGCTAAATATATCTTGGCGAAATTCCTGCGATTGGGATCGT 542

Qy 391 GluPheTyrLysAspIlePhePheAsnAlaLysSerValGlyAspTyrIlePheArgMetTyrPhe 410  
Db 541 GAGTTCACAAAGATTTGGTGAATGCAAAAACATGATGACGATGATGAGACCTTGGAAAT 482  
Qy 411 MetProValHisTyrPheMetValArgHisIleTyrPheProCysLeuArgSerLysIle 430  
Db 481 ATGCGTGCATTAAGGAGGATGCTGCGCAATATTTTCCATGATCTTAAGAAATGCAATA 422  
Qy 431 ProlSerThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCys 450  
Db 421 CCTAAGGAGATTCATGATGATGATCTCTTCTTATATGCTGCTTTCATGATGATGATG 362  
Qy 451 IleAlaValProCysArgLeuPheLysLeuTyrPheAlaPheLysGlyIleMetPheGlnVal 470  
Db 361 ATTCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 302  
Qy 471 ProlSerValPheIlePheHisArgTyrLeuGlnArgPhe---GlySerThrValGlyAsn 489  
Db 301 CCTTGGTCTATCTACAGAACTTCTGCAAAACAGTTCAAAACATGCAATGATGCGCAAC 242  
Qy 490 MetIlePheThrPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyr 509  
Db 241 ATGACATCTCGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 182  
Qy 510 HisAspLeuMetLysArgLysGlySerMetSer 520  
Db 181 CATGATGCTGATGATGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 149

## RESULT 3

LOCUS AY110660 1181 bp mRNA linear HTC 17-OCT-2002  
DEFINITION Zea mays CL739.1 mRNA sequence.  
ACCESSION AY110660  
VERSION AY110660.1 GI:21215250  
KEYWORDS HTC.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.  
1 (bases 1 to 1181)

## REFERENCE

AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitsitt, M.S.,  
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.  
Maize Mapping Project/Dupont Consensus Sequences for Design of  
Overgo Probes  
Unpublished (2002)

## TITLE

JOURNAL 2 (bases 1 to 1181)

## AUTHORS

Cooper, E.H.  
Direct Submission  
Submitted (25-APR-2002) Maize Mapping Project, University of  
Missouri, Columbia, MO 65211, USA  
If you are interested in getting corresponding physical clones,  
these are publicly available from zmdb and may be found by BLAST  
searching at KSL, maizeMap.org; zmdb, www.zmdb.iastate.edu; TIGR,  
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the  
maize cDNA sequences is either Virginia Walbot, Stanford or Pat  
Schubert, Iowa State, then clones may be requested from zmdb:  
www.zmdb.iastate.edu.

## FEATURES

source  
location/Qualifiers  
1..1181  
/organism="Zea mays"  
/mol\_type="mRNA"  
/db\_xref="MaizeDB:632842"  
/db\_xref="taxon:4577"  
/clone\_lib="Maize Mapping Project/Dupont Consensus  
Library"

/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed Dupont contigs; this resource was  
assembled by Dupont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project"

BASE COUNT 322 a 240 c 275 g 327 t 17 others



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 31, 2003, 05:00:54 ; Search time 2515 Seconds  
(without alignments)  
5025.180 Million cell updates/sec

Title: US-09-623-514a-2  
Perfect score: 2771  
Sequence: 1 MAILDSAGVTTYTENGSGGF.....QPMCVLYHYHDMNRKGSMS 520

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+.\_p2n.model -DEV-xlp  
-Q/csq2\_1/USPTO\_spool/US0963511/runat.29082003.152200.62/app.query.fasta.1.711  
-DB-EST -OFMT-fastap -SUFFIX-trst -MINMATCH=0.1 -LOOPT=0 -LOOPEXT=0  
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi LIST=45  
-OCALIGN=200 -THR.SCORE-pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE-LOCAL  
-OUTFMT-pio -NOM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0963511.ecgn.1.1.4237/runat.29082003.152200.62 -NCP=6 -ICPU=3  
-NO\_MMAP -LARGEUDERY -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estp1:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_inv:\*  
19: em\_gss\_p1n:\*  
20: em\_gss\_vrl:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_phg:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1279	46.2	1572	11	AY105372 Zea mays
2	1155	41.7	961	13	B0510367 EST617782
3	1067.5	38.5	1181	11	AY110660 Zea mays
4	1018	36.7	729	12	BI422326 EST532992
5	966.5	34.9	872	14	CB660599 OSJNE021
6	959.5	34.6	813	14	CB651490 OSJNE016H
7	943.5	34.0	851	14	CB660636 OSJNE02J
8	902	32.6	586	13	B0148998 NF086D09F
9	835.5	30.2	741	13	BQ998922 QG620J23
10	832.5	30.0	729	12	BJ322102 BJ322102
11	765	27.6	539	12	BI422212 EST532878
12	762.5	27.5	606	12	B316561 BJ316561
13	734.5	26.5	855	9	AW349274 GM210004B
14	731.5	26.4	696	14	CB619077 OSJTE03G
15	727.5	26.3	540	9	AJ470192 AJ470192
16	726	26.2	572	13	B0124305 EST609881
17	725.5	26.2	820	14	CB619078 OSJTE03G
18	689	24.9	561	9	AW035727 EST281881
19	687.5	24.8	862	10	B6321213 Zm04_0590
20	683.5	24.7	626	9	AV926912 AV926912
21	653	23.6	816	29	B2503532 BONFC89TR
22	652.5	23.5	629	9	AA042298 24635 CD4
23	635.5	22.9	739	14	CB646515 OSJNE009A
24	635.5	22.9	794	14	CB651491 OSJNE016H
25	625.5	22.6	655	13	B0998311 QG619C15
26	600	21.7	673	13	BQ998554 QG619N20
27	598.5	21.6	480	9	AJ470191 AJ470191
28	595.5	21.5	480	9	AJ470190 AJ470190
29	590	21.3	770	14	CB646514 OSJNE009A
30	588	21.2	746	28	BH466416 B0HEK18TR
31	583.5	21.1	736	14	CB620474 OSJTE03G
32	579.5	20.9	828	13	B0861203 QG617N02
33	575.5	20.8	868	14	CB660600 OSJNE021
34	571	20.6	1064	12	BM476315 AGENCOURT
35	562.5	20.3	555	12	BJ268713 BJ268713
36	559	20.2	1043	12	BM919422 AGENCOURT
37	551	19.9	656	13	B0862755 QG621P16
38	546.5	19.7	515	9	AV925760 AV925760
39	542.5	19.6	412	14	CB063593 SAV61E06
40	534.5	19.3	807	14	CB660637 OSJNE02J
41	533	19.2	627	14	CA784021 SAT92d11
42	532.5	19.2	748	14	CB620472 OSJTE03G
43	531.5	19.2	508	13	B0009402 QG610F20
44	529.5	19.1	959	9	AL520752 AL520752
45	528.5	19.1	968	13	B0670343 AGENCOURT

# ALIGNMENTS

RESULT 1	ALIGNMENT	ALIGNMENT
AY105372	AY105372	1572 bp
LOCUS	PC0148220 mRNA sequence.	linear
DEFINITION	AY105372	HTC
ACCESSION	AY105372.1 GI:21208450	HTC
VERSION	AY105372.1	HTC
KEYWORDS	HTC	HTC
SOURCE	Zea mays	HTC
ORGANISM	Zea mays	HTC
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.	HTC
	(bases 1 to 1572)	HTC



```

OY      498  epehcygln  501
      ::::  |||
DB      3015  TTATTCTCAA  3024

Search completed: August 31, 2003, 07:17:10
CPU time : 152 secs

```

```

Db      2238 TTCGACGAGTCGTTTGTATATGAAGGCCACCTATTTGTCCAGAGAGAAAGCTGGCTGG 2297
Oy      268 -----AAlaAapLySaAlaAsn-----ProGluValSerTyrTyrVa 280
Db      2298 TACTAATTACACTAGGAAATCAGACACTGTTCACATACCTACAGTCAACAGTATTT 2357
Oy      280 1SerLeuysSerLeuAlaTyrPheMetValAlaProthrLeuysTyrGlnProserTy 300
Db      2358 G-----TACTCTTAATTTGGCTCTACCCCTTATCTACCGCTGACAGCTA 2399
Oy      300 rProArSerAlaCysIleArGlySerTyrPValAlaArgGlnPheAlaIysLeuValI 320
Db      2400 TCCCAAGAAATCCCACTGTAAATAGGGGTATAGTTGCTATGAAGTTTGCACAGCTTTGG 2459
Oy      320 ePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSe 340
Db      2460 TTGGCTTTTCTATGTGTGTACTACATCTTTGAAGAGCTTTGACCCCTGTGTTCGGAATAT 2519
Oy      340 rLyS-----HisProLeuysGlyAspPheLeuTyrAlaIleGluArgValIleuysLeuSe 359
Db      2520 CAACACAGAGCCCTTCAGCGCT-----GCTGTTCGGTCCATATG 2558
Oy      359 rVal-----ProAsnLeuTyrValTyrPheLeuysMetPheTyrCysPhePh 374
Db      2559 TGTATTTTAACCTCCACTTTGCCAGAGGTGTCGTATCTCTCTCTACTTTTTCGCTTTT 2618
Oy      374 eHisLeuTyrPheAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLy 394
Db      2619 GCATCTGGCTGGCTCAATGCCCTTGTGCTGAGATGTACGCTTTGGTGACAGAGATCTATAT 2678
Oy      394 sAspTyrPheAsnAlaIysSerValGlyAspTyrTyrArgMetThrAsnMetProValH 414
Db      2679 GGATTTGGTGGAGACTCCACGTATACCTCCACACTATATATGAACCTGAATGTGGTGCTCA 2738
Oy      414 sLyAsTyrMetValArgHisIleTyr-----PheProCysLeuArgSerLyIleProLy 432
Db      2739 TGACTGGCTATATTAACATATGCTTTACAGAGACTTTCTCGGTTTTTCTCCAGAGATTCCA 2798
Oy      432 sThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAl 452
Db      2799 ATCTGCTCCATGTTAGCTGTCTTGTGCTATGCTGTACTAGTACAGAAATGTGCTTGGC 2858
Oy      452 sValProCysArg-----LeuPheLysLeuThrAlaPheLeuGlyI 466
Db      2859 TGTG---TGCCTGAGCTTTTCTATCCCGTGCCTTGTGCTGTTCATAGTTCCTTGGAT 2915
Oy      466 eMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerTh 486
Db      2916 GCGCTTTCAC-----TTCATTGTCAATGATATGCGAAAAAG-----CC 2954
Oy      486 rValGlyAsnMetIlePheTyr-----PheIlePheCysAl 498
Db      2955 GATTTGGAATGTCTGTGATGTGACTTCTCTTCTTGGGCAATGAGATCTACTCTGTT 3014
Oy      498 ePheGlyGln 501
Db      3015 TTATTTCTCAA 3024
RESULT 15
US-08-509-187D-2
; Sequence 2, Application US/08509187D
; Patent No. 5834283
; GENERAL INFORMATION:
; APPLICANT: Chang, Ya-Yuan and Chang, Catherine C.Y.
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL ACYLTRANSFERASE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/509,187D
FILING DATE: 31-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lamport Hammitte, Ann
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033cpdv
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4079 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-509-187D-2

Alignment Scores:
Pred. No.: 5.57e-34 Length: 4079
Score: 405.50 Matches: 148
Percent Similarity: 41.36% Conservative: 77
Best Local Similarity: 27.21% Mismatches: 167
Query Match: 14.63% Indels: 153
DB: 2 Gaps: 24

US-09-623-514A-2 (1-520) x US-08-509-187D-2 (1-4079)
QY 17 GlycylglutpHeValAspLeuAspArgLeuArgArgLysSerArgSerAspSer 36
DB 1675 GGCAGTCACTTGTGATGATTTTGTGCACCAATCTATTGAAAGTCAGATCATTTGATTAAT 1734
QY 37 AsnGlyLeu-----LeuLeuSerGlySerAspAsnAsnSerProSer 50
DB 1735 GGTGGGCGCGCTCCACAACTTTTCCTGCTCTTGAGAGAGAAACCAACATAGAGCG 1794
QY 51 AspAspValGlyAlaProAlaAspValArgAspArgIleAspSerValValAsnAspAsp 70
DB 1795 AAGCATTTGAGAGCACTCCAGAA----- 1818
QY 71 AlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyAspAsnAsnGlyGly 90
DB 1819 ---CAAGGAAGATTTTATT----- 1836
QY 91 GlyArgGlyGlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTYRArgProSer 110
DB 1836 ----- 1836
QY 111 ValProAlaHisArgArgAlaArgGluSerProLeuSerSerAspAlaIlePheLysGln 130
DB 1837 -----GCAAGGCGCTCTCTCTTA-----GATGAACGTGTTGAAGTG 1872
QY 131 SerHis--AlaGlyLeuPheAsnLeuGlyValAlaValLeuIleAla---ValAsnSer 148
DB 1873 GACCAACATGACAAACAAATATATACATATGTTATTTCCCTCCATCTCTTATCTACG 1932
QY 149 ArgLeuIleIleGlu-AsnLeuMetLysTYRArgLysPheLeuIleArgThrAspPheTrp 168
DB 1933 ACACCTTGAGTAGATTCATGATGAGGAAGGCGTGGCTGCACAACTTACG----- 1984
QY 168 eSerSerArgSerLeuArgAspTrpProLeuPheMetCysCysAlaIleSerLeuSerIleP 188
DB 1985 -----CCTCGCTGCTTATGAT-----TTGGCAAA-TT 2012

```







```

APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.
TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/509,187D
FILING DATE: 31-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lamport Hamillite, Ann
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033cpdv
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)742-4214
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4011 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1397..3046
US-08-509-187D-3
Alignment Scores:
Pred. No.: 9,16e-35 Length: 4011
Score: 412.50 Matches: 148
Percent Similarity: 42.46% Conservative: 80
Best Local Similarity: 27.56% Mismatches: 173
Query Match: 14.89% Indels: 136
DB: 2 Gaps: 23
US-09-623-514A-2 (1-520) x US-08-509-187D-3 (1-4011)
QY 17 G1yG1yG1uPhueAlaAspIleuAspArgIleuArgArgIysSerArgSerAspSerSer 36
Dd 1604 GGCAGTCACTTGTGATGATTTGTGACCAACATCTATTGAAGAAGTCAGCATCTATTGATAT 1663
QY 37 AsnGlyLeu-----LeuLeuSerGlySerAspAsnAsnSerProSer 50
Dd 1664 GGTGGGCGCCGCTCCACCAACCTTTCTGCTTGTGAAGGAGAGAAAMACCAACATAGACGC 1723
QY 51 AspAspValGlyAlaProAlaAspValArgAspArgIleAspSerValValAsnAspAsp 70
Dd 1724 AAGGATTTGAGACGACCTCCAGAA----- 1747
QY 71 AlAG1nG1yPrrAlaAsnIleuAlG1yAspAsnAsnG1yG1yAspAsnAsnG1yG1y 90
Dd 1748 ---CAGGAAAGATTTTATT----- 1765
QY 91 G1yArgG1yG1yG1yG1yArgG1yAsnAlaAspAlaThrPheThrTyArgProSer 110
Dd 1765 ----- 1765
QY 111 ValProAlaHisArgAlaArgAlaArgGlnSerProLeuSerSerAspAlaIlePheIysGln 130

```

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/121,057  
 FILING DATE:  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:  
 NAME: LAMPORT HAMMITTE, ANN.  
 REGISTRATION NUMBER: 34,858  
 REFERENCE/DOCKET NUMBER: DCI-033CP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (617) 227-2700  
 TELEFAX: (617) 227-5941  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4011 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 FEATURE:  
 NAME/KEY: CDS  
 LOCATION: 1397..3046  
 US-08-121-057-3

Alignment Scores:  
 Pred. No.: 9,16e-35 Length: 4011  
 Score: 412.50 Matches: 148  
 Percent Similarity: 42.46% Conservative: 80  
 Best Local Similarity: 27.56% Mismatches: 173  
 Query Match: 14,898 Indels: 136  
 DB: Gaps: 23

US-09-623-514a-2 (1-520) x US-08-121-057-3 (1-4011)

17 GYGLYGLPHEVALASPLEUASPHARGLEUARGARGYSERASPSER 36  
 1604 GGCAGCTACTTGTGATGATTTTGTGACCATTCATTCATTAAGTACCATCATTTAGTATAT 1663  
 37 ASNGLYLEU-----LEULEUSERYLSERASPSANSERPROSER 50  
 1664 GGTGGGTGGCTGTCACACACTTTTCTGTTCTTGAAGAGAGAAAACCATATACAGG 1723  
 51 ASPAPVALGVALAPROVALASPYVALARGSPATGILEASPSERVALVALASNASPAP 70  
 1724 AAGGATTTGAGAGCACTCCAGAA----- 1747  
 71 ALANGNGLYTHRALASPLEUASLAGLASPASNASNGLYGLYASPASNASNGLYGLY 90  
 1748 ---CAAGCAACATTTTATTT----- 1765  
 91 GLYARGGLYGLYGLUGLYARGGLYASNALASPALATHRPHETHRYRARGPROSER 110  
 1765 ----- 1765  
 111 VALPROLAHISARGARGALARGGLUSERPROLEUSERSERASPALALEPHELYSGLN 130  
 1766 -----GCAAGCGCTCTCTCTTA-----GATCACTGCTTGAAC 1801  
 131 SERHIS---ALAGLYLEUPHEASPLEUCYSVALVALLEUILLLELA---VALANSER 148  
 1802 GACCACTATGACAAATATATATACATGTTATTTGCCCTCTCATTTCTTTATCCACAC 1861  
 149 ARGLEULEILEGLU-ASPLEUAMELYSTYGLYTRPLEUILLLEARGTHRASPHERTRPH 168  
 1862 ACACCTGTGATGATATACATGATGAGAGAGGCTGCTGCTGAGT----- 1907  
 168 ESERASERASERLEUARG-ASPTTRPROLEUPHEMECTSCYSILLESERLEUSER-ILE 187  
 1908 -TCAGCGCTCTCTCTTATGCTTTTGGCAAAATTTCTTACCGTGTGTTGACCTGGGAGAC 1966  
 188 PHEPROLEUALALAPHERTHRVALGLULYSLEUVALLEUGLILYSTYR----- 203  
 1967 ATGTTCTCTCTACATTTTCAAGTTCCCTTATTTCTGTTTCAACATTTGGCGCACTGGCTAT 2026

204 -----LESERGLUPROVALLE-----PHEUHSILLEILLE 216  
 2027 AGCAAGATCTCATCCGCTGATCCGCTCTCTCTTCCATGAGCTTTTTCATGATCTTC 2086  
 217 THRMTHTHGLUVALLEUYYR---PROVALTYRVALTHLEUARGCYSAPSERALAPHE 235  
 2087 CAGATTGAGATTCTAGTTTGGACCAACATATGTTGTGTTA-----GCATAT 2134  
 236 LEUSERGLYVALTHLEUWELLEUETHRYCSILLEVALTRPLEUYSLEUVALSERTYR 255  
 2135 ACACGCCACCAAGCTTCCGGGTTCATATATTTGACAGAGATTCGTTTGTATGATGAG 2194  
 256 ALAHSTHR-----SEPTYRASPILEARGSERLEUALASNALAASPLYSALA 272  
 2195 GCCCATCATTTTGTGACAGAGAACGTCGCTCGGTACTTAATTCACCTAAGAGAAATCA 2254  
 273 ASN-----PROGLUVALSERTYRVALSERLEUYSERLEUALATYRPE 288  
 2255 AGCAGCTGTCCAAATACCTACAGTCAACACAGATATTTG-----TACTTC 2296  
 289 METVALALAPROTHLEUCYSTYRGLNPROSEPTYRPROARGSERALACYSILLEARGYS 308  
 2297 TTATTTGCTCTACCTTATCTTACGCTGACGCTATCCAGAGATCCACTGTAAGATG 2356  
 309 GLYTRPVALLAARGGLNPHLEALALYSLEUVALLEPHERTHGLYPHEMECTGLYPHEILE 328  
 2357 GGTATATGTCGATAGAGAGTTGACAGAGCTTGTGCTTTTCTATGATGATGATGATG 2416  
 329 ILEGLUGLYTRILEASPROILEVALARGANSERLYS---HISPROLEUYSGLYASP 347  
 2417 TTTGAAAGCTTTGTGCCCCCTGTTTCCGAATATCAACAGAGCCCTTACGCCGT--- 2473  
 348 LEULEUVALALILEGLUARGVALLEUYSLEUSERVAL-----PROASN 362  
 2474 -----CGTGTCTGCTCTATGATGATTTTACTCCACTTTCGCCAGGT 2515  
 363 LEUTYRVALTRPLEUCYSMETPHERTYRCSYPHEPHISLEUETHRPLEUASNILLEUALA 382  
 2516 GTGCTGATCTCTCTCTACTTTTGTGCTTTTTCACGCTGCTGCTGCTGCTGCT 2575  
 383 GLULEULEUCYPHEGLYASPARGLUPHETRYRGLYASPTTRIPASNALALYSERVAL 402  
 2576 GAGATGTTACGCTTTGCTGACAGAGATGTTCTATAGAGATTTGGGCAACCTCATATAC 2635  
 403 GLYASPTTRIPARGMETTRIPASMETPROVALHISLYSTRMETVALARGHISILEYR 422  
 2636 TCCAACTATATATAGAACCTGGAATGCTGCTCATGACTGGCTATATTAATGCTTAC 2695  
 423 -----PHEPROCYSLEUARGSERLYSILEPROLYSTRLEUALALEILLEALAPHE 440  
 2696 AAGGACTTCTCTGCTGTTTCTCCAGAGATTCAAATCTGCGGCATGTTAGCTGCTTT 2755  
 441 LEUVALSERALAPHEHISGLULEUCYSILLEALVALPROCYARG----- 456  
 2756 GCTGATCTGCTGATATACAGAAATGCTGCTGCTGCTT-----TCTTGAAGCTTTTCTAT 2812  
 457 -----LEUPLYSLEUETHRPALEUAGLYLLEMEHPHEGLINVALPROLEUVALPHE 474  
 2813 CCCGCTCTTGTGCTCTTCAATGTTCTTGAATGCTTTCAAC-----TTC 2860  
 475 ILETHRASNTRYLEUNGINGLUARGPHEGLYSERTHVALGLYASNMETLEPHERTP--- 493  
 2861 ATTGCAATATGATAGTCGAAAAG-----CCGATTTGGAATGTTCTGATGCTGACT 2911  
 494 -----PHEILEPHECYSILEPHEGLYGLN 501  
 2912 TCTCTTTTCTTGGCAATGAGAGCTTACTGCTTTATTTATTCCAA 2956

RESULT 11  
 US-08-509-187D-3  
 Sequence 3, Application US/08509187D  
 Patent No. 5834283  
 GENERAL INFORMATION:

```

Dh 879 AAG---GAGCATGGAGCTACTGCAGCATGATGAGAGGGCCCTCGAAGCGGGGGTCCCAAT 935
Qy 363 LeuTyrValTlPleuLeuCysMetPheTyrCysPhePheHisLeuTlPleuAsnIleLeuAla 382
Dh 936 CACCTCATCTGGCTGCACATCTTCTTACTGGCTCTTCCACTCTCTGCAATGATGGCCGTGCT 995
Qy 383 GluLeuLeuCysPheGlyAspArgGluPheTyrTyrAspPrtPrtAsnAlaLysSerVal 402
Dh 996 GAGCTCATGCAgTTGGAGACGGGAGTCTTACCGGGAGCTGTGAACTCCGAGTCTGTC 1055
Qy 403 GlyAspTlPrtPrtAspMetPrtPrtAsnMetPrtValHisLysTlPrtMetValArgHisIleTyr 422
Dh 1056 ACCTACTTCTGGCAGAACTGGAACATCCCTGTGCACAAGTGTGTCATCAGCACTTTCAC 1115
Qy 423 PheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuVal 442
Dh 1116 AAGCCCATGCTCTGCACGGGGCAGCAGCAATGAGTGTGGCAGACAGGGGTGTCTTGCC 1175
Qy 443 SerIleValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTlPrtAla 462
Dh 1176 TCGGCGCTTTCACAGAGTACCTGTGTAGCGCTCCCTCTGCAGAAATTCGCGCTTGGGCG 1235
Qy 463 PheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTlPrtLeuGlnGlu 481
Dh 1236 TTTCAGGGGCAATGAGGCTCAGATCCCACTGGCCCTGTGTCTGTGGGGCCGCTTTTTCAGAGGC 1295
Qy 482 ArgPheGlySerThrValGlyAsnMetIlePheTlPrtPheIlePheCysIlePheGlyGln 501
Dh 1296 AACTAT-----GGCAGCAGCACTGTGTGG---CTGTGCTCATCATGCGAGC 1340
Qy 502 PrometCysValLeuLeuTyrTlPrtHisAsp 511
Dh 1341 CCAATGACGCTCTCATGTATGCTTCACGAC 1370

RESULT 9
US-09-103-754A-3
Sequence 3, Application US/09103754A
Patent No. 6344548
GENERAL INFORMATION:
APPLICANT: Farese, Robert
APPLICANT: Casese, Sylvaine
APPLICANT: Smith, Steven
APPLICANT: Erickson, Sandra
TITLE OF INVENTION: Diacylglycerol O-acylttran
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESS: Bozicevic & Reed
STREET: 285 Hamilton Avenue, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,754A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 6510-105P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650 327 3400
TELEFAX: 650 327 3231
TELEX:

```

```

? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
?     LENGTH: 629 base pairs
?     type: nucleic acid
?     STRANDEDNESS: single
?     TOPOLOGY: linear
?     MOLECULE TYPE: cDNA
US-09-103-754A-3

Alignment Scores:
Pred. No.:      1,49e-62          Length:      629
Score:           652.50           Matches:      131
Percent Similarity: 87.58%       Conservative: 3
Best Local Similarity: 85.62%    Mismatches:   10
Query Match:      23.55%         Indels:        9
DB:               Gaps:          1

US-09-623-514A-2 (1-520) x US-09-103-754A-3 (1-629)

QY      304  A l c y s i l e a r g l y s g l y t r p v a l a l a a r g i n p h e a l a l y l s l e u v a l l e p h e r t h g l y      323
Db      2    G C A T G C A T A C G G A A G G G T G G G C T G C T C A A T T G C A A A C T G C A T A T T C A C C G G A      61

QY      324  P h e m e t g l y p h e l l e t l e g l u g i n p r y l l e a s n p r o l l e v a l a r g a s n s e r l y s h i s p r o      343
Db      62  T T C A T G G G A T T T A A T A T A G A C A A T A T A A T C C T A T T G C A G A A C T G A A G C A A C A T C C T      121

QY      344  L e u l y s g l y A s p l e u d e u t y r a l l e g l u a r g v a l l e u l y s l e u s e r v a l p r o a n l e u      363
Db      122  T T G A A A G G G A C T T C T A T A T G C T A T T G A A A G A G T T G A A G C T T T A G T T C C A A T T T A      181

QY      364  T y r v a l t r p l e u c y s m e t h e r t y r c y s p h e p h e i s l e u t r p l e u a s n l l e u a l a g l u      383
Db      182  T A T G T G G C T C G C A T G T T C T A C G C T T C C A C C T T T G G T A A A C A T A T T G G C A G A G      241

QY      384  L e u l e u c y s p h e g l y A s p a r g l u p h e r t y r l y s a s p r t p r p a n a l a l y s s e r v a l g l y      403
Db      242  C T T C T C T G C T T G G G G A T C G T A A T T C T A C A A G A G T T G T G A A T G C A A A A G T G T G G A      301

QY      404  A s p t y r t r p - A r g m e t t r p a s m e t p r o v a l h i s l y s t r p --m e t v a l a r g h i s l e t y r      422
Db      302  G A T T A C T G G G A G A A T G S A A T A T A T G C C G T C C A T A A A G G A T G G G A T G G C C G A C A T A T T A C      361

QY      423  -P h e - P r o c y s l e u a r g s e r l y s l i e -P o l y s t h r l e u a l a l l e l l e l l e a l a p h e l e u      441
Db      362  C T T C C C C G G C T T G C G C A C A A G A A T T A C C C A A G A C A C C C G G C C A T T A A C C A T T G G C T T      421

QY      442  v a l s e r l a -----v a l p h e h i s      447
Db      422  T C C C A A G C C C C C T G G A G G C C T T T C A A T      448

RESULT 10
US-08-121-057-3
Sequence 3, Application US/08121057
Patent No. 5484727
GENERAL INFORMATION:
APPLICANT: CHANG, TA-YUAN
APPLICANT: CHANG, CATHERINE C. Y.
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text

```







Db 423 GCAGGTTATTTTACGAACTTATCATGATGCGCTGCTG--GATCCTATCCAG 479  
 Qy 168 PheSerSerArgSerLeuArgAsp-----TrpProLeuPheMetCysLysLeu 184  
 Db 480 GTGGTCTCTGTTTGAAGACCCCTACAGCTGGCTCCCATCCGATGATTATGCA 539  
 Qy 185 LeuSerLlePheProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIle 204  
 Db 540 TCCAAATTTTGTGTGGCTGCATTTTCAGATTGAGAAAGCGCTGGCAGATGGGCTG 599  
 Qy 205 SerGluProValIlePheLeuHisLleLleLlePheThrGluValLeuTyrPro 224  
 Db 600 ACAGAGAGATGGGGCTGCTGCTACCTGCTTACCTGGCCACAAATCTTGGCTTCCA 659  
 Qy 225 ValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeu 244  
 Db 660 GCAGCTGTGGCTTACGTGAGTGTATCATCCATCCAGTGGGTCCGTTGCTTGCGCA 719  
 Qy 245 Thr---CysLleValTyrLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg 263  
 Db 720 TCATACTCCATCATGTTCTCCATCAAGCTTATTCCTACCGGATGTCACACTGTGGTGGC 779  
 Qy 264 -----SerLeuAlaAlaAlaAsp 270  
 Db 780 CAGCGAAGGTCAGAGCCAAAGCTGCTCTACAGGAGGAGGAGGCTGGGCTGCT-- 836  
 Qy 271 LysAlaAsnProGluValSerTyr-----TyrValSerLeuLysSerLeuAlaTyrPhe 288  
 Db 837 ---GCCAGAGACGTGAGCTATCCAGAACCTGACCTACGAGATCTCTATCTACTTC 893  
 Qy 289 MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysLleArgLys 308  
 Db 894 ATCTTGCTCTCTACTTGTGTATGAACTCAACTTCTGCTGGTCCCGCAATACGAAAG 953  
 Qy 309 GlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle 328  
 Db 954 CGCTTCTGCTACAGACAGTCTTGGATGCTCTTTTACCCAGCTTCAAGTGGGGCTG 1013  
 Qy 329 IleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu 348  
 Db 1014 ATCCAAAGAGATGGAGTCTTACATCCACAACTCCATGAAAGCCCTTCAG--GATATG 1070  
 Qy 349 LeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsnLeuTyrValTyr 366  
 Db 1071 GACTATTACAGGATCATATGAGCGCTCTTAAAGCTGGCGTCCCAACCATCTGATCTGG 1130  
 Qy 367 LeuCysMetPheTyrCysPhePheHisLeuTyrPheAsnIleLeuAlaGluLeuLeuCys 386  
 Db 1131 CTATATCTTCTTATTTGTTTCCACTCTGCTGCTCATGCTGTGGCAGAGCTTCTGAC 1190  
 Qy 387 PheGlyAspArgGluPheTyrLysAspTyrTrpAsnAlaLysSerValGlyAspTyrTrp 406  
 Db 1191 TTGGAGACCCGCGAGTCTTACAGAGATTTGGATGCTGAGTCTGCTCCTACTTTGG 1250  
 Qy 407 ArgMetTrpAsnMetProValHisLysTyrMetValArgHisLleTyrPheProCysLeu 426  
 Db 1251 CAGAACTGGAATATCCCGCTGACAACTGTGATCAGACACTTCTCAAGAGCTATCTCTC 1310  
 Qy 427 ArgSerLysLleProLysThrLeuAlaLleLleLleAlaPheLeuValSerAlaValPhe 446  
 Db 1311 AGACATGCGCAGACCAATGGGTGGCCAGACAGAGATATTTTGGACCTTACCTTCTTC 1370  
 Qy 447 HisGluLeuCysLleAlaValProCysArgLeuPheLysLeuTyrPalapheLysGlyIle 466  
 Db 1371 CAGTAGACCTAGTAGAGCGTTCCTCCCTGCGATGTCCGCTGTGGGATTCACAGCATG 1430  
 Qy 467 MetPheGlnValProLeuValPheLleThrAsnTyrLeuGlnGluArgPheGlySerThr 486  
 Db 1431 ATGGCTTAGGTCCACAGCGCTGATGTGGCGCATTC-----TTCCAGAGGAGAC 1481  
 Qy 487 ValGlyAsnMetLlePheTrpPheLlePheCysLlePheGlyGlnProMetCysValLeu 506  
 Db 1482 TATGGCAATGACAGCTGTGTGG---GTGACACTCATCATTTGGGCAACCGGTGGCTGTCTC 1538

Qy 507 LeuTyrTyrHisAsp 511  
 Db 1539 ATGATGTCACAGAC 1553  
 RESULT 6  
 US-09-326-203A-15  
 : Sequence 15, Application US/09326203A  
 : Patent No. 6444876  
 : GENERAL INFORMATION:  
 : APPLICANT: Lassar, Mike  
 : APPLICANT: Ruzinsky, Diane  
 : TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic  
 : FILE REFERENCE: 17045/00/NO  
 : CURRENT APPLICATION NUMBER: US/09/326, 203A  
 : PRIOR APPLICATION NUMBER: 60/088, 143  
 : PRIOR FILING DATE: 1998-06-05  
 : PRIOR APPLICATION NUMBER: 60/108, 389  
 : PRIOR FILING DATE: 1998-11-12  
 : NUMBER OF SEQ. ID NOS.: 46  
 : SOFTWARE: PatentIn Ver. 2.1  
 : SEQ. ID NO. 15  
 : LENGTH: 1766  
 : TYPE: DNA  
 : ORGANISM: Rattus sp.  
 US-09-326-203A-15  
 Alignment Scores:  
 Pred. No.: 3,79e-75 Length: 1766  
 Score: 773.50 Matches: 193  
 Percent Similarity: 53.58% Conservative: 84  
 Best Local Similarity: 37.33% Mismatches: 171  
 Query Match: 27.91% Indels: 69  
 DB: 4 Gaps: 17  
 US-09-623-514a-2 (1-520) x US-09-326-203a-15 (1-1766)  
 Qy 14 GluAsnGlyGlyGluPheValAspLeuAspArgLeuArgArgLysSerArgSer 33  
 Db 99 CAGGCTGTGATGGGCCCATGCTAGACGAAGAGAGGTGGCAGACCCGCTGTGGCCCC 158  
 Qy 34 AspSerSerAsnGlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspAspVal 53  
 Db 159 GAC-----TTGGGGCGCGGGGGGTGACCTCGCGTCCGGCT--CCGGTT 200  
 Qy 54 GlyAlaProAlaAspValArgAspArgLleAspSerValAlaAsnAspAlaGlnGly 73  
 Db 201 CCGGCTCCACCCACACACCCGCGACAA-----GACCGGCGAGACC 239  
 Qy 74 ThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyAspAsnAsnGlyGlyArgGly 93  
 Db 240 AGCGTC-----GGCGAGGC 254  
 Qy 94 GlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSerValProAla 113  
 Db 255 CACTGGAGCTGAGG-----TGC 272  
 Qy 114 HisArgArgAlaArgGluSerProLeuSerSerAspAlaLlePheLysGlnSerHisAla 133  
 Db 273 CAT---CGTTCGACAGACTTGTGTCAGCTCAGACAGCGGTTTC--ACCAATTACCT 326  
 Qy 134 GlyLeuPheAsnLeuCysValValIleLleAlaValAlaAsnSerArgLeuLleGlu 153  
 Db 327 GGTATCTGATTTGGTGGCTGGTGTATCTCATCTGATGATGCAAGGTTATTTTAGAG 386  
 Qy 154 AsnLeuMetLysTyrGlyTyrLeuLleArgThrAspPheTrpPheSerArgSerLeu 173  
 Db 387 AATCTTATCAAGATAGCATCTGTGGT--GATCCCATCCAGAGGTGTCTGTTCTG 443  
 Qy 174 ArgAsp-----TrpProLeuPheMetCysLysSerLeuSerLlePheProLeu 190

```

Db      820 GGGCTCCCTGCTGGCGCTGATGGCGACACACATCTTCTCCACAGCTCTCTCTACCG 879
Qy      256 ahsthrserTyrAspIleargSer-----LeuAlaAsnAl 268
Db      880 C-----GAGCTCAACTCATGTGTGCCGACAGGCCAGGCCACAGGCTGCTGCG 927
Qy      268 aaIaaAspIysAlaAsnPro-----GluValSerTyr-----TyrValSe 281
Db      928 AGGGAAAGAGCCAGAGTGTCTGCTCCCGGACACCGTGAAGCTTACCCGACAAATGTGAC 987
Qy      281 rleuYsserLeuAlaTyrPheMetValAlaProThrLeuCYsTyrGlnProSerTyrP 301
Db      988 CTACCGCGAATCTTACTACTCTCTCTCCGCCCGCCACTGTGTGTAGAGCTCAACTTCC 1047
Qy      301 oArgSerAlaCYsIleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuValIlePh 321
Db      1048 CCGCTCTCCCGCTCCGAGAGGCGCTTGTGCTGCGACAGGATCTTGAGATGCTGTCTT 1107
Qy      321 eThrgLysPheMetGlyPheIleIleGlnIleGlnIleAsnProIleValAlaArgAsnSer 341
Db      1108 CACCCAGCTCCAGAGTGGGCTGATCCAGAGTGAATGTCCCGCACCATCCAGAACTCCAT 1167
Qy      341 sHisProLeuLysGlyAspLeuLeuTyrAla-----IleGluArgValLeuLysLeuSe 359
Db      1168 GAAGCCCTTCAG--GACATGAGCTACTACACCATCATCGACGCGCTCTCGAAGCTGGC 1224
Qy      359 rValProAsnLeuTyrValTyrPLeuCYsMetPheTyrCYsPhePheHisLeuTyrLeuAs 379
Db      1225 GGTCCCAATTCACCTCTGCTGCTCATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGAA 1284
Qy      379 nIleLeuAlaGluLeuLeuCYsPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAl 399
Db      1285 TGCCGTGGTGAAGCTATGATGAGAACCGGGAGTCTACCGGGAGCTGGTGAACATC 1344
Qy      399 alySerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArg 419
Db      1345 CAGGTGTCTCAGCTACTTCTGTGAGAGAACTGAGAACTCCTGTGACAAAGTGGTGCATCAG 1404
Qy      419 gHisIleTyrPheProCYsLeuArgSerLysIleProLysTrpHisAlaIleIleIleAl 439
Db      1405 ACACCTCTACAAAGCCATGCTTCGACGGGGACAGCACAGTGAATGCCAGACAGAGGGGT 1464
Qy      439 aPheLeuValSerAlaValPheHisGluLeuCYsIleAlaValProCYsArgLeuPheLys 459
Db      1465 GTTCTGGGCTGGCTTCTTCTTCACAGAACTGTGAGAGCTCTCTCGAATGTTCGG 1524
Qy      459 sLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTyr 478
Db      1525 CCTGTGGCTTTCACGGGAGATATGCTCAGATCCACACTGGCTGTGTGTGGCCGCTT 1584
Qy      478 rLeuGlnGluArgPheGlySerThrValGlyLysAsnMetIlePheTrpPheIlePheCYsIl 498
Db      1585 TTTCCAGGGCAGAACTAT-----GGCAACGAGAGTGTGTG---CTGCGGCTCAT 1629
Qy      498 ePheGlyGlnProMetCYsValLeuLeuTyrTyrHisAsp 511
Db      1630 CATCGGACAGCAATAGCCGCTCCTCATGTACGTCCAGAC 1669

```

RESULT 5  
 US-09-103-754A-2  
 Sequence 2, Application US/09103754A  
 Patent No. 6344548

GENERAL INFORMATION:  
 APPLICANT: Farese, Robert  
 APPLICANT: Cases, Sylvaine  
 APPLICANT: Smith, Steven  
 APPLICANT: Erickson, Sandra  
 TITLE OF INVENTION: Diacylglycerol O-acyltan  
 NUMBER OF SEQUENCES: 6  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Bozicevic & Reed  
 STREET: 285 Hamilton Avenue, Suite 200

```

? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94301
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/103,754A
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Field, Bret E
? REGISTRATION NUMBER: 37,620
? REFERENCE/DOCKET NUMBER: 6510-105P
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650 327 3400
? TELEFAX: 650 327 3231
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1650 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-103-754A-2

Alignment Scores:
Pred. No.: 3,39e-75 Length: 1650
Score: 773.50 Matches: 191
Percent Similarity: 51.818 Conservative: 81
Best Local Similarity: 36.388 Mismatches: 178
Query Match: 27,918 Indels: 75
DB: 4 Gaps: 15

US-09-623-514A-2 (1-520) x US-09-103-754A-2 (1-1650)
Qy      8 GlyValThrThrValThrGluAsnGlyGlyGluPheValAspLeuAspArgLeuArg 27
Db      141 GGCTCGGGGTTTCCGTCCAGAGGTGTAGTGGCCCAAGTGAAGAGCAGAGGTGCGCA 200
Qy      28 ArgArgLysSerArgSerAspSerAsnGlyLeuLeuLeuSerCylSerAspAsnAsn 47
Db      201 GACGCGGCTGTGAGCCCGAC-----TTGGGCGCGCGGGGTGACGCGCGC 245
Qy      48 SerProSerAspAspValGlyAlaProAlaAspValArgAspArgIleAspSerVal 67
Db      246 GCTCGCGCTCCG-----GCTCCAGCCCATACCCGGGACAAA----- 281
Qy      68 AsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyAspAsn 87
Db      281 ----- 281
Qy      88 AsnGlyGlyGlyArgGlyGlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyr 107
Db      282 ---GACGGCGGACACAGCTGGGCGCGCTCTACTGGATCTGAGGTGC----- 326
Qy      108 ArgProSerValProAlaHisArgAlaArgGluSerProLeuSerSerAspAlaIle 127
Db      327 -----CATGCTGTGCAAGATTTCTTTGTTCAAGCTCAAGACAGTGTGT 365
Qy      128 PheLysGlnSerHisAlaGlyLeuPheAsnLeuCYsValAlaValLeuIleAlaValAsn 147
Db      366 TTC---AGCAATATAGTGTATCTCGAATGTGTGTGTGTGTGTATGATCTGATAT 422
Qy      148 SerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrPLeuIleArgThrAspPheTrp 167

```

```

Db 181 CAGTTATTTACACCATGACCGAGGCTTGTATTCACGTATATGTCACCTAAGGTGAT 240
Qy 233 SerAlaPheLeuSerGlyValThrLeuMetLeuThrCysIleValITrPLeuLysLeu 252
    |||||
Db 241 TCTGCTTCTTATCAGGTGACACCTGATGCTGCTCTCTCATGCTATGCTGCTCAAGTGG 300
Qy 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaIleAspLysAla 272
    |||||
Db 301 GTTCTTACCTCCTCACTACTACTATGACATAGAACCTCACTAATTCATGATGATAAGGCC 360
Qy 273 AspProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetValAlaPro 292
    |||||
Db 361 AATCTGTAAGTCTCTCTACTATGTTAGTGAAGGCTTGGCTATTTTCATGCTGCTGCC 420
Qy 293 ThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTyrValAla 312
    |||||
Db 421 ACATGGTATTATGACCAAGCTATCCACGTTCTCCATGTATCCGGAAGGGTGGGTGGCT 480
Qy 313 ArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIleGlnIleTyr 332
    |||||
Db 481 CGTCAATTTCCAAACGTCATATTCACGAGCTCATGGGATTTATATAGACCAATAT 540
Qy 333 IleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLysLeuTyrAlaIle 352
    |||||
Db 541 ATTAATCTCTATTGTAGAACTCAAGACATCTCTGAAAGGGGACCTGTATATGCTATT 600
Qy 353 GluArgValLeuLysLeuSerValProAsnLeuTyrValITrPLeuCysMetPheTyrCys 372
    |||||
Db 601 GAAAGAGCTGTGAAGCTTCAGTTCCAAATCTATATGTTGGCTCCTCAGTTCTACTGCG 660
Qy 373 PhePheHisLeuTrpLeuAlaIleLeuAlaGluLeuLysCysPheGlyAspArgGlnPhe 392
    |||||
Db 661 TTCTCCACCTTTGGTTAAACATATTTGGCAGAGCTCTGCTGCGGGGACCGGAATTC 720
Qy 393 TyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetPro 412
    |||||
Db 721 TACAAATATGTTGGATGCAAAAGCGTTGGAGATTATTTGGAAATGTGAAATATGCT 780
Qy 413 ValHisLysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIleProLys 432
    |||||
Db 781 GTTCACAAATGATGGTTCACATGTATCTTCCGCTGCGCACAATACGATACCAAAA 840
Qy 433 ThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLysCysIleAla 452
    |||||
Db 841 GTACCCGCAATTCATGCTTCTTATGCTGTGCACTCTTTCATGAGTATGATGCA 900
Qy 453 ValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeu 472
    |||||
Db 901 GTTCCTTGCGGCTCTCTCATATGATGCGCTTTCATGGAATATATGTTCAAGTCCCTTG 960
Qy 473 ValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePhe 492
    |||||
Db 961 GTCCTTATCAAAACCTTTTACAAAGAGTTGGCTCCATGCGGGAACATGATCTTT 1020
Qy 493 TrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAspLeu 512
    |||||
Db 1021 GGTTCACGCTTCTGCAATTTGGAGCAACGATGTGGCTCTTATATACCAAGACCTG 1080
Qy 513 MetAsnArgLysGlySerMetSer 520
    |||||
Db 1081 ATGACCGCAAAAGATCATGTCC 1104

```

```

; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1976
; TYPE: DNA
; ORGANISM: Yeast
US-09-165-042-2

Alignment Scores:
Pred. No.: 2,51e-77 Length: 1976
Score: 794.00 Matches: 205
Percent Similarity: 52.53% Conservative: 86
Best Local Similarity: 37.00% Mismatches: 190
Query Match: 28.65% Indels: 73
Gaps: 20

US-09-623-514a-2 (1-520) x US-09-165-042-2 (1-1976)
Qy 4 LeuAspSerAlaGly-----ValThrThrValThrGluAsnGlyGlyGluPheVal 21
    |||||
Db 91 GTTGAACGCGGTGTGTGAGCGGCTCACCCGGGCTACGGGCGGCGGAGGGGAGATGGCGG 150
Qy 22 AspLeuAspArgLeuArg-----ArgLysSerArgSerAspSerSer 36
    |||||
Db 151 CCGTTGCTTAGGGGCCGAGGAGTGGGGCGCGCGCTCGGGGCTACGACACCGGACAGCC 210
Qy 37 AsnGlyLeuLeuSerGlySer----- 44
    |||||
Db 211 CACGCTTGCGTGGCGGCGGCTGCGGCTGAGGCCATGAGCCAGCCGCGGACCTCCGCGG 270
Qy 45 -----Asp-AsnAsnSerProSerAsp---AspValGlyAlaProAlaAspValArgAs 61
    |||||
Db 271 CCGAGAGACAGAGGTGCGGCGCTCGAGCACAGCGGCGGCGGCGGCTCGGCGGCGGAGA 330
Qy 61 PArgIleAspSerValValAsnAspArgIleGly-----ThrAlaAsnLeuAlaGlyAs 80
    |||||
Db 331 AGAG-----GTGCGGAGCGCGCTGCGGCGGCGGCGGAGCTGGAGCGGCGGGA 378
Qy 80 PAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGluArgGlyAs 100
    |||||
Db 379 CGGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
Qy 100 nAlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgArgAlaArgGluSe 120
    |||||
Db 439 GCTGAGCTGC-----CATGCGCTCGCAGATTC 465
Qy 120 rProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVa 140
    |||||
Db 466 TTTATTGAGCTGTGACAGTGGCTTC---ACCACTACCGTGGCACTCTGAACTGGTGT 522
Qy 140 lValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTr 160
    |||||
Db 523 GGTATCTATATCTTGAGCAATGCCGCTTATTTCTGGAAGACTCATCAAGATGCGAT 582
Qy 160 PleuIleArgThrAspPheTrpPheSerArgSerLeuArgAsp-----TrpPr 177
    |||||
Db 583 CTTGGTG---GACCCCAATCCAGAGTGTCTCTCTCTCAAGAGATCCCATAGCTGGCC 639
Qy 177 OleuPheMetCysLysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGlu 197
    |||||
Db 640 CGGCCCATGCTGCTGTTATTCGGCCAAATGCTTGTGCTGCTCATTCACAGTGGAGA 699
Qy 197 sLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeuHisIleIleThr 217
    |||||
Db 700 GCGCTGCGGCTGGGTGCTGCTGACGAGCAGGCGGCGGACTGCTGCTGACGTAGCAACT 759
Qy 217 rMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSe 237
    |||||
Db 760 GGCACCATTCCTGTTCCAGCGGCTGTGCTTCTACTGTTGAGTCACTCACTCAGT 819
Qy 237 rGlyValThrLeuMetLeuThr---CysIleValITrPLeuLysLeuValSerTyrAl 256
    |||||

```

```

; RESULT 4
; US-09-165-042-2
; Sequence 2, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
; APPLICANT: Sturley, Stephen L.
; APPLICANT: Oelkers, Peter
; TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
; FILE REFERENCE: 0575/56331
; CURRENT APPLICATION NUMBER: US/09/165,042

```

Db 112 CTC-----TCCGATACTTCCCGTCGGACGATGTTGGACCTCGCGCCCGCA 159  
 Qy 60 ArgAspArgIleAspSerValValAsnAspAlaIngIlyThrAlaAsnLeuAlaIy 79  
 Db 160 AGGAGATCGGGTGGATTCGCTCGCGAGAGAGGCTCAGAGGAACAGCAATTAGCT 216  
 Qy 80 AspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGly 99  
 Db 217 -----GGCGGAGATCGCCGAACTAGGGAATCCGCGGAGCG----- 252  
 Qy 100 AsnAlaAspAlaThrPheThrTyArgProSerValProAlaHisArgAlaArgIu 119  
 Db 253 -----GATGTAAGGTTTACGATACGACCTCGCTCCAGCTCATCGAGAGAGGAG 306  
 Qy 120 SerProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeu 139  
 Db 307 ACTCCCTCAGCTCGACGCTATCTTCAAAACAAAGCATGAGGATTTGTTCAACCTCTG 366  
 Qy 140 ValValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyGly 159  
 Db 367 GTAGTGTGTTCTGTTGCTGTAAAGTAACTATCATGAAACCTCATGAAGTATGGT 426  
 Qy 160 TrpLeuIleArgThrAspPheTrpPheSerArgSerLeuArgAspTrpProLeuPhe 179  
 Db 427 TGGTTGATCAGAACTGATTTTGGTTTACTTCTACATCTTACAGAGACTGGCCGCTTTC 486  
 Qy 180 MetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuVal 199  
 Db 487 ATGTGTGTTCTTCACTTTCGCTGCTTCTTGGCTGCTTCAACGCTTCAAGAAATGGTA 546  
 Qy 200 LeuGlnLysTyTrIleSerGluProValValIlePheLeuHisIleIleIlePheMetThr 219  
 Db 547 CTTCGAAATTCATATCTGAGCGCTGTGCAATCTTTCATGCTTAAACCAAGACA 606  
 Qy 220 GluValLeuTyProValTyValThrLeuArgCysAspSerAlaPheLeuSerGlyVal 239  
 Db 607 GAGGCTCTGTATCAGCTACCTCAGCTCAGCTGATGCTGCTTCTTCTGAGGTGTC 666  
 Qy 240 ThrLeuMetLeuLeuThrCysIleValTyLeuLysLeuValSerTyAlaHisThrSer 259  
 Db 667 AGTTGATGCTGCTCAGCTGATGCTGCTGAGATGGTGGTCTTACCTCATCTACTAGC 726  
 Qy 260 TyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyTr 279  
 Db 727 TAGCAGATTAAGAACCTCGCCAAATCTGATTAAGTGAATCCGAAATCTCCACTAT 786  
 Qy 280 ValSerLeuLysSerLeuAlaTyTrPheMetValAlaProThrLeuLysTyGlnProSer 299  
 Db 787 GTTACTTCAAGACCTGGCGTATTTCAATGCTTCCCACTGCTGTTATACGCCAACG 846  
 Qy 300 TyrProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuVal 319  
 Db 847 TATCCAGCTTCTCATATGATCCGAAAGGTTGGTGGCTCGTCAACTGCAAAAGCTGTC 906  
 Qy 320 IlePheThrGlyPheMetGlyPheIleIleGluGlnTyTrIleAsnProIleValArgAsn 339  
 Db 907 AATATCACAGCACTAGGATTTAATAAGACATATATATATATATATATATATATAT 966  
 Qy 340 SerLysHisProLeuLysGluAspLeuLeuTyAlaIleGluArgValLeuLysLeuSer 359  
 Db 967 TCMAAGCATCTCTGTAAGGAGACCTTCTATATGCTATGAAAGATGTTGAAGCTTTTA 1026  
 Qy 360 ValProAsnLeuTyValTrpLeuLysMetPheTyCysPhePheHisLeuTrpLeuAsn 379  
 Db 1027 GTTCAAAATCTATATGTTGGTGTGATGCTTCTTCTTCTTCTTCTTCTTCTTCTT 1086  
 Qy 380 IleLeuAlaGluLeuLeuLysPheGlyAspArgGluPheTyTrLysAspTrpTrpAsnAla 399  
 Db 1087 AATATGGCAGAGCTCTCTGCTGCGGAGCCTGAATCTTCAAAAGATGTTGGAATGCA 1146  
 Qy 400 LysSerValGlyAspTyTrTrpArgMetTrpAsnMetProValHisLysTrpMetValArg 419  
 Db 1147 AAAAGCGTTGGAGATTATTTGAGAAATGGGAATATGCTTCAAAATGATGTTTCA 1206

Qy 420 HisIleTyTrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAla 439  
 Db 1207 CATGTAATCTTCCGTCGCTCGGCATCAAGATACAAAAGTACCCTTATCATTTAGCT 1266  
 Qy 440 PheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLys 459  
 Db 1267 TTTCTTAGTCTCGCACTCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1326  
 Qy 460 LeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyTrLeu 479  
 Db 1327 CTAATGGCTTTCATGGAATTAATGTTTCAAGTCCCTTGGTCTTATACAAACTTTTA 1386  
 Qy 480 GlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePhe 499  
 Db 1387 CAAGAAAGGTTGGCTGCATGAGGGGGAACATGATCTTGGTCACTTCTTGCATTTTC 1446  
 Qy 500 GlyGlnProMetCysValLeuLeuTyTrHisAspLeuMetAsnArgLysGlySerMet 519  
 Db 1447 GGACAAACGATGTGTGGCTTCTTATATACATGATGATGATGATGATGATGATGATG 1506  
 Qy 520 Ser 520  
 Db 1507 TCC 1509  
 Db 1507 TCC 1509  
 RESULT 3  
 US-09-593-359-1  
 ; Sequence 1, Application US/09593359  
 ; Patent No. 6552250  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Laroche, Andre J.  
 ; APPLICANT: Nykiforuk, Cory L.  
 ; APPLICANT: Weslake, Randall J.  
 ; TITLE OF INVENTION: Diacylglycerol O-acyltransferase  
 ; FILE REFERENCE: 24015US0  
 ; CURRENT APPLICATION NUMBER: US/09/593,359  
 ; CURRENT FILING DATE: 2000-06-14  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 1  
 ; LENGTH: 1446  
 ; TYPE: DNA  
 ; ORGANISM: Brassica napus  
 ; FEATURE:  
 ; OTHER INFORMATION: Dgat2  
 ; NAME/KEY: CDS  
 ; LOCATION: (82)..(1107)  
 US-09-593-359-1  
 Alignment Scores:  
 Pred. No.: 4,43e-192 Length: 1446  
 Score: 1833.00 Matches: 342  
 Percent Similarity: 95.65 Conservative: 10  
 Best Local Similarity: 92.938 Mismatches: 16  
 Query Match: 66.158 Indels: 1  
 Gaps: 0  
 US-09-623-514A-2 (1-520) x US-09-593-359-1 (1-1446)  
 Qy 153 GluAsnLeuMetLysTyGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSer 172  
 Db 2 GAATAATCTCATGAGTACGTTGTTGATGATGATGATGATGATGATGATGATGATGAT 61  
 Qy 173 LeuArgAspTrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAla 192  
 Db 62 CTGCGAGATTG-CGCGTTTTCATGTTGTCTCTCCCTTTCATCTTTCCTTGGCTGCC 120  
 Qy 193 PheThrValGluLysLeuValLeuGlnLysTyTrIleSerGluProValValIlePheLeu 212  
 Db 121 TTACCGTGAAGAAATTAATCTTCAAGAAATGATATCTGAACTTGTGTCATCTTCTT 180  
 Qy 213 HisIleIleIleThrMetThrGluValLeuTyTrProValTyTrValThrLeuArgCysAsp 232

```

Db 237 ATGGCATTGTTGGATTCTGCTGCGCTTACTACGGTGAACGAGAACGGTGGCGAGAGTTC 296
QY 21 ValAspLeuAspArgLeuArgArgArgSerArgSerArgSerArgSerArgSerArgSerArg 40
Db 297 GTGAGCTTGTGATGGCTGCTGCGACGGAATGAGATTCGATCTTCTTAACGAGACTTCTT 356
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspPvalGlyAlaProAlaAspValArg 60
Db 357 CTCTCGGTTCCCATATAATTCCTCGGATGATGTTGAGACTCCCGCGAGAGTTAG 416
QY 61 AspArgIleAspSerValAlaAsnAspAspAlaGlyIleThrAlaAsnLeuAlaGlyAsp 80
Db 417 GATCGGATGATTCCTGTTAAGAGAGCTGACGAGGAGACACCAATTTGGCCGGAGAT 476
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGly 100
Db 477 AATAACGGTGGTGGCGATATAACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 536
QY 101 AlaAspAlaThrPheThrTyraArgProSerValProAlaHisArgArgAlaArgGlySer 120
Db 537 GCCGATGCTACGTTTACGTACGACCGTGGTCCGCTCAGCTCATCGAGGGCGAGAGAGT 596
QY 121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
Db 597 CCAGTACGCTCCGACGCAATCTTCAACAGAGCCATGCGGATTTCAACCTCTGTGTA 656
QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGlyAsnLeuMetIleTyrglyTrp 160
Db 657 GTAGTTCTTATCTGCTGTAACAGTACAGCTCATCATGAAATTTATGAAAGTATGGTGG 716
QY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
Db 717 TTGATCAGAACGAGATTCGTGGTTAGTTCAGATCCGTCGAGAAATGGCCGCTTTCATG 776
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGlyLeuValLeu 200
Db 777 TGTGTATATCCCTTTCGATCTTCTTCTTGGCTGCTTTCAGGTTGAGAAATGGTACCT 836
QY 201 GlnIleTyrlleSerGlyProValIlePheLeuHisIleIleIleIleIleIleIleIle 220
Db 837 CAGAAATACATATCAGAACCTGTGTCATCTTCTCATATATATATATATATATATATAT 896
QY 221 ValLeuTrpProValTyraIleThrLeuArgCysAspSerIleAlaPheLeuSerGlyVal 240
Db 897 GTTTGTATCCAGTTTACGTCACCTAAGGTGATTCGCTTTTATATAGAGTGCATCT 956
QY 241 LeuMetLeuLeuThrCysIleValIlePheLeuValSerTyraIleHisIleTrpSerTy 260
Db 957 TTGATGCTCTCTCATCTTGCATGTTGTGCTAAAGTTGGTTTCTTATGCTCATACTAGCT 1016
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspIleAlaAsnProGlyValSerTyTyraVal 280
Db 1017 GACATTAAGATCCCTACGCAATGACGATGATTAAGCCCAATCTCTAAGCTCTCTAGCT 1076
QY 281 SerLeuIleSerLeuAlaTyrrPheMetValAlaProThrLeuGlyIleProSerTy 300
Db 1077 AGCTTAAAGGCTTGGCATATTTTCATGCTGCTCCACCATTTGCTTATTCGCAAGTAT 1136
QY 301 ProArgSerAlaCysIleArgGlyGlyTrpValAlaArgGlnPheAlaValIleValIle 320
Db 1137 CCACGTTTCGATGATACGGAAGGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1196
QY 321 PheThrGlyPheMetGlyPheIleIleGlyGlnTyrlleAsnProIleValAlaGlnSer 340
Db 1197 TTCACGGATTCATGGGATTTATATAGAACATATATATATATATATATATATATATAT 1256
QY 341 LysHisProLeuGlyGlyAspLeuLeuTyraIleIleGlyValLeuIleValSerVal 360
Db 1257 AACCATCTTGTGAAGCGATCTTCTATATATGTAAGAGTGTGAAGCTTTCAGT 1316
QY 361 ProAsnLeuTyraIleTrpLeuGlyMetPheTyrrCysPhePheHisLeuTrpLeuAsnIle 380
Db 1317 CCAAAATTTATATGTGTGCTGTGCTGTCTACTGCTTCTTCCACCTTTGGTTAAACATA 1376

```

```

QY 381 LeuAlaGlyLeuLeuGlyCysPheGlyAspArgGlyPheTyrrLysAspTrpTrpAsnAlaLys 400
Db 1377 TTGCAAGAGCTTCTCTGCTCGGGGATTCGGAATTCACAAATATGTGTGATGCAAAA 1436
QY 401 SerValGlyAspTyrrTrpArgMetTrpAsnMetProValHisIleTrpMetValArgHis 420
Db 1437 AGTGTGGAGATTTACGCAAGATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 1496
QY 421 IleTyrrPheProCysLeuArgSerIleIleProIleThrLeuAlaIleIleIleAlaPhe 440
Db 1497 ATATACCTCCCGCTGCTTCCGACAGATATCCAAAGACACTGCGCATATACATGCTTTC 1556
QY 441 LeuValSerAlaValPheHisGlyLeuGlyIleAlaValProCysArgLeuPheGlyLeu 460
Db 1557 CTAGTCTCTGCACTCTTTCATGAGCATGATGATGATGATGATGATGATGATGATGATGAT 1616
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrrLeuGln 480
Db 1617 TGGCTTTTCTTGGATATATGTTACAGTCCCTTGTGCTTCATCAACAACTATCTACAG 1676
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
Db 1677 GAAAGGTTTGGCTCAACGCTGGGGAACATGATCTTGTGTCATCTTGTGATTTGCGA 1736
QY 501 GlnProMetCysValLeuLeuTyrrTyrrHisAspLeuMetAsnArgLysGlySerMetSer 520
Db 1737 CAACCGATGTGTGTGCTTCTTATATACACAGACTGATGAAACGGAAGATGATGATGCA 1796

RESULT 2
US-09-593-359-3
; Sequence 3, Application US/09593359
; Patent No. 6552250
; GENERAL INFORMATION:
; APPLICANT: Larcoche, Andre J.
; APPLICANT: Nykiforuk, Cory L.
; APPLICANT: Weselake, Randall J.
; TITLE OF INVENTION: Diacylglycerol O-acyltransferase
; FILE REFERENCE: 24015050
; CURRENT APPLICATION NUMBER: US/09/593,359
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; OTHER INFORMATION: DGAT1
; NAME/KEY: CDS
; LOCATION: (1)..(1512)
US-09-593-359-3

Alignment Scores:
Pred. No.: 1.43e-244 Length: 1512
Score: 2309.50 Matches: 442
Percent Similarity: 88.10% Conservative: 17
Best Local Similarity: 83.84% Mismatches: 43
Query Match: 83.35% Indels: 19
DB: 4 Gaps: 5

US-09-623-514a-2 (1-520) x US-09-593-359-3 (1-1512)
QY 1 MetAlaIleLeuAspSerAlaGlyAlaThrThr---ValThrGlnAsnGlyGlyGlyGlu 19
Db 1 ATGGCATTGTTGGATTCTGAGGCGTGCCTGACCCGCGAGAGAACGGC-----51
QY 20 PheValAspLeuAspArgLeuArgArgGlySerArgSerArgSerArgSerArgSerArg 39
Db 52 GTGCGGATCTCGACAGGCTCCACCGTCGTAATTCGAGTTCGATTCCTTCCACGAGACTC 111
QY 40 LeuLeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspVal 59
Db 111

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 31, 2003, 05:06:49 ; Search time 97 Seconds  
(without alignments)  
2366.178 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771  
Sequence: 1 MAILDSAGVTTVTENGGEF.....OPMCVLLYYHLMNKRKGSMS 520

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+p2n.model -DEV=xlp  
-O=/cg2.1/USPTO.spool/US09623514/runat.29082003.152200.102/app.query.fasta.1.711  
-DB=Issued\_Patents\_NA -QEXT=fastap -SUFFIX=rm1 -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09623514 @CGN.1.1.85.ernaat.29082003.152200.102 -NCPD=6 -ICPD=3  
-NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAP=10 -XGAPEXT=0.5 -FGAP=6  
-FGAPEXT=7 -YGAP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued\_Patents\_NA:\*  
1: /cg2.6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cg2.6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cg2.6/ptodata/1/ina/6A.COMB.seq:\*  
4: /cg2.6/ptodata/1/ina/6B.COMB.seq:\*  
5: /cg2.6/ptodata/1/ina/PTUS.COMB.seq:\*  
6: /cg2.6/ptodata/1/ina/backfiles1.seq:\*

-Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2771	100.0	1942	4	US-09-326-203A-1
2	2309.5	83.3	1512	4	US-09-593-359-3
3	1833	66.1	1446	4	US-09-593-359-1
4	794	28.7	1976	3	US-09-165-042-2
5	773.5	27.9	1650	4	US-09-103-754A-2
6	773.5	27.9	1766	4	US-09-326-203A-15
7	773.5	27.9	1766	4	US-09-326-203A-16
8	709.5	25.6	1895	4	US-09-326-203A-14
9	652.5	22.9	629	4	US-09-103-754A-3
10	412.5	14.9	4011	1	US-08-121-057-3
11	412.5	14.9	4011	2	US-08-509-187D-3
12	412.5	14.9	4011	2	US-09-121-396-3

13	412.5	14.9	4011	5	PCT-US93-09704A-3	Sequence 3, Appl1
14	405.5	14.6	4079	1	US-08-121-057-2	Sequence 2, Appl1
15	405.5	14.6	4079	2	US-08-509-187D-2	Sequence 2, Appl1
16	405.5	14.6	4079	5	US-09-121-396-2	Sequence 2, Appl1
17	405.5	14.6	4079	5	PCT-US93-09704A-2	Sequence 2, Appl1
18	375	13.5	1607	4	US-09-328-857A-1	Sequence 1, Appl1
19	364.5	13.2	1509	4	US-09-328-857A-2	Sequence 2, Appl1
20	364	13.1	2040	3	US-09-165-042-4	Sequence 4, Appl1
21	299	10.8	275	4	US-09-326-203A-5	Sequence 4, Appl1
22	284.5	10.3	253	4	US-09-326-203A-7	Sequence 7, Appl1
23	284	10.2	234	4	US-09-326-203A-3	Sequence 3, Appl1
24	277	10.0	267	4	US-09-326-203A-8	Sequence 8, Appl1
25	275	9.9	254	4	US-09-326-203A-4	Sequence 4, Appl1
26	275	9.9	254	4	US-09-313-294A-580	Sequence 580, App
27	266	9.6	262	4	US-09-326-203A-9	Sequence 9, Appl1
28	266	9.6	262	4	US-09-313-294A-1662	Sequence 1662, Ap
29	224.5	8.1	518	4	US-09-326-203A-12	Sequence 12, Appl
30	187.5	6.8	519	4	US-09-326-203A-11	Sequence 11, Appl
31	166.5	6.0	325	4	US-09-326-203A-10	Sequence 10, Appl
32	151	5.4	996	1	US-08-121-057-1	Sequence 1, Appl1
33	151	5.4	996	2	US-08-509-187D-1	Sequence 1, Appl1
34	151	5.4	996	2	US-09-121-396-1	Sequence 1, Appl1
35	151	5.4	996	5	PCT-US93-09704A-1	Sequence 1, Appl1
36	150	5.4	6755	3	US-08-931-999-4	Sequence 1, Appl1
37	144	5.2	1239	4	US-09-134-001C-2048	Sequence 2048, Ap
38	142.5	5.1	3967	4	US-09-672-749-14	Sequence 4, Appl1
39	122.5	4.4	1230025	4	US-09-198-452A-1	Sequence 1, Appl1
40	120	4.3	3707	1	US-08-118-101A-1	Sequence 1, Appl1
41	117	4.2	7557	1	US-08-464-136-4	Sequence 4, Appl1
42	117	4.2	7557	2	US-08-349-131-4	Sequence 4, Appl1
43	117	4.2	7557	3	US-08-470-297A-4	Sequence 4, Appl1
44	117	4.2	7557	5	PCT-US91-07149-4	Sequence 4, Appl1
45	114	4.1	1029	4	US-09-205-815B-45	Sequence 45, Appl1

#### ALIGNMENTS

RESULT 1  
US-09-326-203A-1  
; Sequence 1, Application US/09326203A  
; Patent No. 6444876  
; GENERAL INFORMATION:  
; APPLICANT: Ruzinsky, Mike  
; TITLE OF INVENTION: Acid-CoA:Cholesterol Acyltransferase Related Nucleic  
; FILE REFERENCE: 17045/00/MO  
; CURRENT FILING DATE: 1999-06-04  
; PRIOR APPLICATION NUMBER: 60/088,143  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/108,389  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: Patent Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 1942  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-326-203A-1

#### Alignment Scores:

Pred. No.: 2,916-295  
Score: 2771.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 4  
Length: 1942  
Matches: 520  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-623-514A-2 (1-520) x US-09-326-203A-1 (1-1942)

QY 1 MetalalieleuaspSerAlaGlyValThrThrValIThrGluAsnGlyGlyGluPhe 20  
|||||





```

DB      1275 CGGAGTTTATAGAGATTGGTGGAAATGCGAGACGTGGAGAGTACTGGAGAAATGTGG 1334
QY      410 AsnMetProValHisIstYrMetValAGHISileTyrPheProCysLeuArgSerLys 429
DB      1335 AATATCCGTCATGAATGATGATGTTGGCATATATCTGTCATGCTTACAAAATGGA 1394
QY      430 IleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeu 449
DB      1395 ATACCAAAAGATAGTGCAGATTGATCGCGTTCTTGTGTGTCGATTTTTCATGAGCTG 1454
QY      450 CysIleAlaValProCysArgLeuPheLysLeuThrPheAlaPheLeuGlyIleMetPheGln 469
DB      1455 TCGGTGACGATCCCTTGGCAAAATTCAGATTGTTGGCGTCTCGGGTATCATAGCTTCAG 1514
QY      470 ValProLeuValPheIleThrAsnTyrLeuGlnGluArgPhe---GlySerThrValGly 488
DB      1515 GTTCCTCTGTAATGCTGACTAATTAATCTTGCAAGAAAAGTTCAAAAATCAATGATGGTGGC 1574
QY      489 AsnMetIlePheThrPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyr 508
DB      1575 AATATGATGTCTGCGTCTTCTGTATCTTGTGGCAACCTATGATGTGTGTGTGTATC 1634
QY      509 TyrHisAspLeuMetAsnArgLysGlySer 518
DB      1635 TACACGACTGTATGATGAAAGAGCTACT 1664

RESULT 15
AF129003 2099 bp mRNA linear PLN 22-DEC-1999
LOCUS Nicotiana tabacum diacylglycerol acylCoA acyltransferase (DAGAT)
DEFINITION mRNA, complete cds.
ACCESSION AF129003
VERSION AF129003.1 GI:6625652
KEYWORDS
SOURCE Nicotiana tabacum (common tobacco)
ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 2099)
Bouvier-Nave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and
Schaller, H.
Expression in yeast and tobacco of plant cDNAs encoding acyl
CoA:diacylglycerol acyltransferase
Eur. J. Biochem. 267 (1), 85-96 (2000)
20069349
10601854
2 (bases 1 to 2099)
Benveniste, P.
Direct Submission
Submitted (17-FEB-1999) Plant Molecular Biology Institute,
C.N.R.S., 28 rue Goethe, Strasbourg 67083, France

FEATURES
Source
1..2099
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="xanthi SH6"
/db_xref="taxon:4097"
/tissue_type="calli derived from leaf protoplasts"
1..2099
/gene="DAGAT"
/gene="DAGAT"
208..1806
/function="catalyzes the acylation of diacylglycerol by
oleoylCoA to give triacylglycerol"
/note="similar to acyl-CoA cholesterol acyltransferase"
/codon_start=1
/product="diacylglycerol acylCoA acyltransferase"
/protein_id="AAFI9345.1"
/db_xref="GI:6625653"
/translation="NMIMELPESVEMKTTTTSGLIENLNSDLNHSVRRRSGNGEPA
SAINSSDANMSERDVCSSGAGLETVERSKSVGESSDVIRKEDRDNDNVANGESK

```

```

STTTTTTPKRAYRASAPAHRIKESPLSSDAIFKQSHAGLENLCVVVLIAVNSRLII
ENIMYGLIIRAGFEFSSKSLDMDPLMLCISLIQLPLAFLVERLAQORHLERAVY
TLMTITTYALIIYPLVILIGDSAFELIYILMLVACIYMKLVASYAHTHOKRLAKS
TKDKESDSDSFYDSFKSLAYFMAVPTCYOLSLSPHPCIKRGVARGFTLVIETG
LMGRTIEOYINPIYONSOPLKGNILAIERIKLVSYPULYWLMEFCFPHILNII
AEILCEGDEEFTKDMNNAKTTIDEIRNMMNMPHAKMVRITFPCLNGLPKVALILIA
FLVSAVFELICIAVPCRLERKMAFMKIMFQVPLVLTNLOKRFQSSVMGNMFWCFE
CIIQPMCLLYHDHVMNRKSSAR"

BASE COUNT      543 a      405 c      485 g      666 t
ORIGIN
Alignment Scores:
Pred. No.:      2,39e-144      Length:      2099
Score:          1842.50      Matches:      361
Percent Similarity: 76.82%      Conservative: 50
Best Local Similarity: 67.48%      Mismatches: 100
Query Match:    66.49%      Indels:      25
DB:              8      Gaps:      7

US-09-623-514a-2 (1-520) x AF129003 (1-2099)
QY      1 MetAlaIleLeuAspSerIaGlyValIMrThrValThrcIuAsnGlyGlyGlyuphe 20
DB      217 ATGGAATTCGCCGAGAGCGTCGAAATGACGACGACGACGACGACGATTCGGGTATGGAAC 276
QY      21 ValAspLeuAsp-----ArgLeuAgaArgLysSerArgSerAspSerAsn 37
DB      277 CTCACCTCCGATCTTAATCAGTCGCTGTCGAGAGACG-----GGCAGTAAAT 324
QY      38 GlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAla 57
DB      325 GGTTTGAGCGCGTAGTCGCAATTAACAGTTG-----GATCGCAATGTGAGCGAA 375
QY      58 AspValArgAspArgLysAspSer-----ValValAsnAspAspAlaGln 72
DB      376 GATGAGAGATGTGTGTGCGACGCGTCTGATTCGAAACGATGATGAGCGAGTAA 435
QY      73 GlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyArg 92
DB      436 TCGGTTCGTCAGTCACGATGATGT-AATTCGAAAGGAGGACGACGACGAAATGATGCTGC 494
QY      92 GelyGlyGlyGlyArgGlyAsnAlaAspAlaThr-----PheThrTyrAr 108
DB      495 GATGTGAGGAGGAAAGCAATCAACGGAACAACGACGCGCTTAATTTAGCTTACAG 554
QY      108 gProSerValProAlaHisArgArgAlaArgGlySerProLeuSerSerAspAlaIlePh 128
DB      555 GCGTCGCGACCAAGCTCACCGCGGAATCAAGAGAGTCTCTCAGCTCCGACGCCATTYT 614
QY      128 elysGlnSerHisAlaGlyLeuPheAsnLeuCysValValIleAlaValAsnSe 148
DB      615 CAACAGATTCACGCGAGCGCTTCAATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 674
QY      148 ArgLeuIleIleGluAsnLeuMetLysTyrGlyThrPheIleArgThrAspPheThrPh 168
DB      675 CAGGCTGATTCAGGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 734
QY      168 eSerSerArgSerLeuArgAspThrProLeuPheMetCysIleSerLeuSerIlePh 188
DB      735 TACCTGAGAGTCGTGAGGATGCGCTTCAATGTCGTCTCAGCTCCAAATTTT 794
QY      188 eProLeuAlaAlaPheThrValGlyLysLeuValLeuGlnLysTyrIleSerGluProVa 208
DB      795 GCGCGTCGCTGCTTTCTTGTGTGAGAGATGGCAGACGACGACGACGATTCACGACGCTGC 854
QY      208 lValIlePheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrValTh 228
DB      855 GGTGTACTCTTCATACATACATTAACAACACGCTGATTTGTGTGATTCGATTCGATCAT 914
QY      228 rLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleVa 248
DB      915 TCTTGAGTGTATTCCTCTTCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 974

```

RESULT 14  
AF298815  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
TITLE  
SUBMISSION  
Submitted (23-AUG-2000) Division of Biochemistry, NIASR, RDA, 249  
Seodun-dong, Gwonsun-gu, Suwon 441-707, Korea

1964 bp mRNA linear PLN 16-OCT-2000  
Perilla frutescens diacylglycerol acyltransferase (Dgat1) mRNA,  
complete cds.  
AF298815  
AF298815.1 GI:10803052

Perilla frutescens  
Perilla frutescens  
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:  
Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:  
Asterales: Lamiales: Lamiales: Nepentaceae: Nepentaceae: Elsholtziaceae;  
Perilla.

1 (bases 1 to 1964)  
Hwang, S.-K. and Hwang, Y.-S.  
Isolation of Perilla frutescens diacylglycerol acyltransferase cDNA  
Unpublished  
2 (bases 1 to 1964)  
Hwang, S.-K. and Hwang, Y.-S.  
Direct Submission  
Submitted (23-AUG-2000) Division of Biochemistry, NIASR, RDA, 249  
Seodun-dong, Gwonsun-gu, Suwon 441-707, Korea

FEATURES  
source  
1..1964  
/organism="Perilla frutescens"  
/mol\_type="mRNA"  
/cultivar="Okdong"  
/db\_xref="taxon:48386"  
1..1964  
/gene="Dgat1"  
69..1673  
/gene="Dgat1"  
/note="Dgat1"  
/product="diacylglycerol acyltransferase"  
/protein\_id="AA233696.1"  
/db\_xref="GI:10803053"  
/translation="MALDSPETLDTSSADNGAAHHTLRROARSVPPLDSDS  
NSLSEASINDSENANDANLLENLGGAVESNEKOESEKGEKAVKNGESEN  
GTEVAVKFTFEPAPAHNRKESPLSDAIFKOSHAGLENCIVLVAVNSLLEN  
LMKGYLKSFGFSTSLRDPMLMCCLSLPVALASPELVELVNTYIPENAVSL  
HTTITVELLFPYVVLIRKDSANLSCVTTMLLFRCTYWKAKIVSAHNTYDIRAKSLD  
KWEASRWMLDADYDSEKSLAVFMAVPLCYPSYPTACIRKGVAVOLIKLVIF  
TGLMGRITIDYINPQVONSHPLKMLVAIEERVLKLSVNLVYVLCMECFPHLMTN  
IALELLCFIDREFYKDMNNARVVEEYRMNNMNVHKKMYHICPCIONGIPKIVAVL  
IAELVSAIFHELCVAVPCOIFKEMAFSGIMLVPLVTVNYLDEKFKNSVMGMEMFC  
FCIPGOMPVLIVYHDMNRKASAR"

BASE COUNT 522 a 395 c 461 g 586 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,39e-145 Length: 1964  
Score: 1854.00 Matches: 359  
Percent Similarity: 74.18% Conservative: 49  
Best Local Similarity: 65.27% Mismatches: 92  
Query Match: 66.91% Indels: 50  
DB: 8 Gaps: 9

US-09-623-514a-2 (1-520) x AF298815 (1-1964)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrThrValThr-----GluAsnGly 16  
DB 69 ATGGCGATCTTGGACCTCGCGAGATCCCGACACGACGATCGACAGCAACAGGC 128  
QY 17 GlyGlyGluPheValAspLeuAspArgLeuArgArgGlySerArgSerArgSer 36  
DB 129 GCCGACATCAACCACTCTT-----CGCCGAGAGAAAGTGCAGCTCGTCCGCC 182  
QY 37 AsnGlyLeuLeuSerGlySerAspAsnSerProSerAspValGlyAlaPro 56  
DB 183 -----CTTTCGACCTCGGATCCCACTCTCTG----- 209

QY 57 AlaAspValArgAspArgIleAspSerValAlaAsnAsp-----AspAlaGlnGlyThr 74  
DB 210 -----GAGCAGAGAGCGCAATCAATGATTCGAAATGTTCCAAACGAC 254  
QY 75 AlaAsnLeuAlaGlyAspAsnAsnGlyGly----- 84  
DB 255 GCTAATTGATCGAAATCTCCGCGGCGGAGCCGTGGAATCCGAGAACGAAACAGAC 314  
QY 85 -----GlyAspAsnAsnGly-----GlyGlyArgGlyGlyGlu 96  
DB 315 AGTTATGTTAGAGAGAGAGAGGCGGAAAGTACAGAGAAATGCAAACTAGTATGCGCAC 374  
QY 97 GlyArgGlyAsnAlaAspAlaThrPheThrTyrrArgProSerValProAlaHisArgArg 116  
DB 375 GGAACGTATGTTATGCGCGCTCAATTCACATTCAGCGCGGCGGCGCTGCTCAGCCGAA 434  
QY 117 AlaArgGluSerProLeuSerSerAspAlaIlePheLeuGlnSerHisAlaGlyLeuPhe 136  
DB 435 AATAGAGAGATCCTCTTACCTCCAGCCCATCTTCAAACAGAGCATGACGCGCTCTTC 494  
QY 137 AsnLeuGlyValValIleLeuAlaValAsnSerArgGluLeuIleGluAsnLeuMet 156  
DB 495 AACCTTGTATGATGAGGCTGCTGCTGCTGTAATAGCAGACTAATATGGAATTTAATG 554  
QY 157 LysTyrrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrp 176  
DB 555 AAGTATGGGTGGCTGATCAATCAGCATGTTGGTTTATGTTCAACATCGCTTAGGATGG 614  
QY 177 ProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGlu 196  
DB 615 CCACTGCTAATGTTGTTGCTTACGCTTCACGTTTTCACATTCCTCATTTCTGTCGAC 674  
QY 197 LysLeuValLeuGlnLysTyrrIleSerGluProValIlePheLeuHisIleIle 216  
DB 675 AAGTTGGTGAACATAATATATACAGCATGGGTGCGCATCTTCTCATGTTACATAC 734  
QY 217 ThrMetThrGluValLeuTyrrProValTyrrValThrLeuArgCysAspSerAlaPheLeu 236  
DB 735 ACAACAGTGGAAATCTGTTCCAGTGTGTGATCTTATGTTGTTGTTGTTGTTGTTGTT 794  
QY 237 SerGluValThrLeuMetLeuLeuThrCysIleValIlePheLeuLysLeuValSerTyrrAla 256  
DB 795 TCAAGGTCTACGCTAATGCTTCTTCTGCTGACGTGATAGTTGTAAGCTCGTTCTCAGCA 854  
QY 257 HisThrSerTyrrAspIleAspSerLeuAlaAsnAlaAspLys----- 271  
DB 855 CATACAACTATGATTTGAGAGACTGTCGAAATCACTTATATAGCGGAAAGCTATGTC 914  
QY 272 -----AlaAsnProGluValSerTyrrTyrrValSerLeuLysSerLeuAlaTyrrPheMet 289  
DB 915 AGGTACTGGAACCTCGACCTACGCTTATGATGTAAGTTAAGTCTGGCATCTTCATG 974  
QY 290 ValAlaProThrLeuGlyTyrrGlnProSerTyrrProArgSerAlaCysIleArgGly 309  
DB 975 GTTGCTCTCAATGTTGTTACAGCAAGCTACCTCGCAGCTTCGATTCGGAAGGCT 1034  
QY 310 TrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIle 329  
DB 1035 TGGGTGTAAGGCACTAATTAAGCTGTAATATTCACAGACTCATGGATTTATTATA 1094  
QY 330 GluGlnTyrrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu 349  
DB 1095 GAACAGATCACTAATACCGATCGTTCAAAATTCACATCTCTGAAAGAAACCTTTA 1154  
QY 350 TyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrrValTrpLeuCysMet 369  
DB 1155 TATGCCATTGAGAGGCTTGAAGCTTCGTTCCAAATTTAATATGTTGCTCTCATG 1214  
QY 370 PheTyrrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysCysPheGlyAsp 389  
DB 1215 TTTTATGTTTTCACACCTGCTGCTGTAATATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1274  
QY 390 ArgGluPheTyrrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrrTrpArgMetTrp 409

AUTHORS Mielkiewska, E., Pedersen, K., Katavlic, V. and Taylor, D.C.  
 TITLE Direct Submission  
 JOURNAL Submitted (12-MAR-2002) Seed Oil Biotechnology, National Research  
 Council of Canada, Plant Biotechnology Institute, 110 Gymnasium  
 Place, Saskatoon, SK S7N 0W9, Canada

FEATURES  
 source Location/Qualifiers  
 1.2090

/organism="Tropaeolum majus"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:4020"  
 171.1727  
 /codon\_start=1  
 /product="putative diacylglycerol acyltransferase"  
 /protein\_id="AA03340.1"  
 /db\_xref="GI:20086324"

CDS  
 /translation="MAVAESSQNTTMSGSDSDINFRRRKSSVIEPSSSGFTST  
 NGVPATGHAENRQDQVAGAMENATGVSVLGNGGVVGNENKQVEDIDETGSPS  
 PFAHRVRESPLSDAIFRQSHAGLFNLCIVLIANSLLTIENLMKWMGLDTCRWF  
 NSRSGDSIFKMCITLPRFLAFLIVKLVORNNHSELVAVLHVYSTAVLTPVI  
 VLTCDVYMSGVMLFECIMWLKLVSTAHSSDRIKLAKGVAGDAHPNSTIVCS  
 YDVSLSLAIFWVAPTLGYQPSYPRSSCIKGMVVOFVKLVIFGLMGIIEQYINP  
 IYRNSKHPLKGDFLYIERVLYLVNLYVMCMFYSFHLMLNIAELLRGDEEY  
 KDMNAKTYAEYKMMNMPVHRMVRHLTFPCLRNGIPREGAIIITAFIVSGAFHELCI  
 APCHVFELMARIGIMFOVPLVLTINYLQEKRSNMGMTWIFLCILIGRMCVLLLY  
 YNDLINKER"

BASE COUNT 527 a 397 c 458 g 708 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 1 74e-147 Length: 2090  
 Score: 1880.00 Matches: 370  
 Percent Similarity: 78.11% Conservative: 44  
 Best Local Similarity: 69.81% Mismatches: 88  
 Query Match: 67.85% Indels: 28  
 DB: 8 Gaps: 10

US-09-623-514a-2 (1-520) x AT084052 (1-2090)

QY 1 MetAlaIleLeuAspSerIleaglyValThrThrValThrGluAsnGlyGlyGluPhe 20  
 DB 171 ATGGCGGGTGGCAGAGTCGCACGAGAACGACGACA---ACCATGAGTGGTCACGGCGAC--- 224  
 QY 21 ValAspLeuAspArgLeuArgArgArgGlySerArgSerAsp-----SerSer 36  
 DB 225 TCGGATCTCCACAAATTTCCGTAGAGAGAAACGAGATTCCTCCGTGATTGACCTTCGTG 284  
 QY 37 AsnGlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaPro 56  
 DB 285 TCCGGTTTACA-----TCCACCAATGGCGTACCGGCG-----ACTGGCCACGCG 329  
 QY 57 AlaAspValArgAspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsn 76  
 DB 330 GGTGAGAAATCGTACGACGATCGGAGTATGAGGCGTATGAGACGACAAACAGATCGGTCAAC 389  
 QY 77 LeuAlaGlyAspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlu 96  
 DB 390 TTAATATGGA-----AATGGGAGGCGGTGTTATCGGGAATGAA 428  
 QY 97 GlyArg-----GlyAsnAlaAspAlaThrPheThrValArgProSerValProAlaHis 114  
 DB 429 GAGAAACAGTAGGGGACACTATACATTCACCTTACCGGCTTCGTTCCGGCTCAT 488  
 QY 115 ArgArgAlaArgGlnSerProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGly 134  
 DB 489 CGAGAGGTGAGGAGAGTCTCTTACCTCTGATGCAATCTTCAACAGAGCCATGGGGT 548  
 QY 135 LeuPheAsnLeuGlyValValLeuIleAlaValAsnSerArgLeuIleIleGluAsn 154  
 DB 549 TTATTCACCTTGATAGTATGATGCTCATTCAGTAAACAGTAGGCTTATCATTCGAAAT 608  
 QY 155 LeuMetValSerGlyTyrLeuIleArgThrAspPheTyrPheSerSerArgSerLeuArg 174  
 DB 609 CTATATGAACTATGTTGTTGATTCGATCTGTTGTTGTTTAACTCAAGATCATCGGCT 668

QY 175 AspTyrProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaIlePheThr 194  
 DB 669 GATGGTCATCTTTATGCTGTGCTTACACTCCCAATTTCCCACTTCCTGCTTTAT 728  
 QY 195 ValGlyLeuLeuValLeuIleGlySerGlyIleSerGluProValIlePheLeuHisIle 214  
 DB 729 GTTGAAGACGTGGTGGAGCAATCATATATGTAAGCTTGTCTCTCTCATGTA 788  
 QY 215 IleIlePheMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAla 234  
 DB 789 ATCGTTTCAACCGCGCATTTTATATCCAGTATTTGTATCTTAAACGTGTATTCGGTG 848  
 QY 235 PheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTyrPheLeuValSer 254  
 DB 849 TATAGTCTGGTGTGATGATGATGCTTGTGTTGATCATATATGTTGTAAGCTGGTGTA 908  
 QY 255 TyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLys-----Ala 272  
 DB 909 TATGCACATCATGTTCTGATATATAGAACACGCGCAAACTGGCTATTAAGGGGATGCG 968  
 QY 273 AsnPro-----GluValSerTyrTyrValSerLeuLysSerLeuAlaTyr 287  
 DB 969 CACCCCAATTCACCACTTGTGAGTTCCTCATATGATGTTAGCTTGAAGCTTTGGCATAC 1028  
 QY 288 PheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArg 307  
 DB 1029 TTCATGTGGCGCGGACATTAATGTTACGACGCTAGTATCTCGTTCGTCGTATCCGC 1088  
 QY 308 LysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPhe 327  
 DB 1089 AAGGGTGGTGTGTTGCTGCAATTTGCAAACTAATGTTTATAGACTCATGGGGCTTC 1148  
 QY 328 IleIleGlnGlyTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAsp 347  
 DB 1149 ATTATAGAACATATATTTATCTTATCTTCTGCAATTTCCAAACCCCATGGAAGAGAT 1208  
 QY 348 LeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTyrLeu 367  
 DB 1209 TTTTATATGACATAGAAAGAGTTTGAAGCTTTCAGTTCCAAATCATATATGTTGGCTT 1268  
 QY 368 CysMetPheTyrCysPhePheHisLeuTyrPheAsnIleLeuAlaGlyLeuLeuCysPhe 387  
 DB 1269 TGCATGTTTACTCTTTTTCACCTCGTGTGGAACATCTGCTGACGCTTCTTCGCTTT 1328  
 QY 388 GlyAspArgGlnPheTyrLysAspTyrProAsnAlaLysSerValGlyAspTyrTyrArg 407  
 DB 1329 GGTGATCGTGAATTCACAAAGATTGTGGAATGCAAAACCTGTTCGGAGATTTGGAAA 1388  
 QY 408 MetTyrAsnMetProValHisLysTyrMetValArgHisIleTyrPheProCysLeuArg 427  
 DB 1389 ATGTGGAATATGCTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1448  
 QY 428 SerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHis 447  
 DB 1449 AATGGATATCCCAAGGAAGAGTGCATTAATTCGCGCTTGTAGTTTCGCTCTTCAT 1508  
 QY 448 GluLeuCysIleAlaValProCysArgLeuPheLysLeuTyrAlaPheLeuGlyIleMet 467  
 DB 1509 GAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1568  
 QY 468 PheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPhe---GlySerThr 486  
 DB 1569 TTTCAGTTCCTCTGTTATGATTAACGATTAATCAAGAAAGTTCAGTAACTCTATG 1628  
 QY 487 ValGlyAsnMetIlePheThrPheIlePheCysIlePheGlyGlnProMetCysValLeu 506  
 DB 1629 GTGGGCAATATGATCTTCTGTTCACTTCTGATCACTTCTGATCACTTCTGATCACTTCT 1688  
 QY 507 LeuTyrTyrHisAspLeuMetAsnArgLys 516  
 DB 1689 CTATATTAACATGACCTGATTAATCTAAG 1718

Db	91958	ACAGGGGTATCTGCTTTTATGAGGTGTCACCTTGATGCCCTCCACTGATGTGT	92017	Db	93036	CCTTCCTTTGTACACAAATAGAAAGAGAGCTAATTAAGATCTTGTTCCTTGACAG	93095
Qy	249	rpleuylsleuValSerTyrAlaHisThrSerTyrAspIleArgSerIleuAlaAsnAla	269	Qy	412	ProValHisTyrPheMetValArgHisIleTyrPheProCysLeuArgSerIlePro	431
Db	92018	GGCTAAAGTGGTGTCTTATGCTCATACACTAGCATTAAGATCCCTTAGCCAAATGACG	92077	Db	93096	CCGTTCATTAATGATGGTGGTCCACATATATACTTCCCGTGGTGGCCACAGATACCA	93155
Qy	269	laAspLysAla	272	Qy	432	Lys	432
Db	92078	CTGATTAAGGT-AAAATACGAAAAAGAGCGTATGTATGATCATTGCACCTGTGTTACTG	92136	Db	93156	AA-GGTGAGTACATATATACGATATGCAATTTGTGAGATTTGTTTCTGTATATAAT	93214
Qy	273	-----AsnProGluValSerTyrTyrValSerI	282	Qy	433	-----ThrLeuAlaIleIleAlaPheLeuValSer	443
Db	92137	TTTAAACCAACACTGTATGAACCTTAGGCCAATCTCGAAGCTCCTACTAGTACGT	92196	Db	93215	TAAACCTCCACACACTGTTTTCAGACACTGCCATATATATCTTCTTCCAGTCTCT	93274
Qy	282	eulysSerIleuAlaTyrPheMetValAlaProThrLeuCysTyr	296	Qy	444	AlaValPheHisGlu	448
Db	92197	TGAAGAGCTTGGCATATTCATGCTGCTGCCACATGTGTATTCAGTACGTCAAAAGT	92256	Db	93275	GCAGTCTTCATAGAGTATACATACCTTTCATATGCGCCGTCTCTAGACGCATGAAC	93334
Qy	296	-----	296	Qy	448	-----	448
Db	92257	GCATCAACCATTTCTTACTTGCAGAGTTTCTGTCTAAACCTCGAATCTTGTCTTTC	92316	Db	93335	ACGCTAGTGAAGAAATGCTAATATTCAAAGCATGTGTTTACTTAACGATCTTGTTTA	93394
Qy	297	--GlnProSerTyrProArgSerAlaCysIleArgLysGlyTTPValAlaArgGlnPhe	316	Qy	449	-----LeuCysIleAlaValProCysArgLeuPheLysLeuTTPala	462
Db	92317	CCAGCCAAAGTATCCACGCTTCTGCATGATACGAGAGGTTGGTGGCTCGTCAATTTG	92376	Db	93395	CAAAATTCCTTTGGACAGCTATGCAATCGAGTTCCTGTCTCTTCAAGCTATGGCT	93454
Qy	316	laLysLeuValIlePheThrGlyPheMetGlyPheIleIleGlu	330	Qy	463	PheLeuGlyIleMetPheGln	469
Db	92377	CAAAATGTGTCATATTCACCGGATTCATGGATTTATATAGAACAGTACGTTTCACA	92436	Db	93455	TTTCTTGGGATTAATGTTTCAGTTAAATAATTAACTAAGTCTGCAGCTGATTTTACTA	93514
Qy	330	-----	330	Qy	470	-----ValProLeuValPheIle	475
Db	92437	TCTTGCTTATTAGTTTCTCCGTGGTGAATAATCATCCCTGCTGTACACACTGACT	92496	Db	93515	AACTTAATCTCATATTCGACCAACCAATTTGTTGAGTAGTGCTTGGTCTTCAAC	93574
Qy	331	-----GlnTyrIleAsnProIleValArgAsnSerLysHis	342	Qy	476	ThrAsnTyrLeuGlnIleValArgPheLysSerThrVal	487
Db	92497	TCATGTTCTTTTGTATACATTTGGCAGATATATAATCTATGTCTCAGAACTCAAGCAT	92556	Db	93575	ACAACTATCTACAGAAAGTGTGGTCAACGGT-ATGCTCTCAAAACCCGAGAAATA	93633
Qy	343	ProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsn	362	Qy	487	-----	487
Db	92557	CCTTGAAGCGGATTTCTATATGCTATTTGAAGAGTGTGAAGCTTCAGTTCCAAT	92616	Db	93634	GAAGCAATTAACCTTCTTTCATACGCTAGCCATTTAATTCGCAATGCTGAACCTTAATA	93693
Qy	363	LeuTyrValTyrLeuCysMetPheTyrCysPhePheHisLeuTyr	377	Qy	488	-----GlyAsnMetIlePheTyr	493
Db	92617	TTATATGTGTGGCTGCATGTGTCTACTGCTTCTCCACCTTGGATGCTGTGATCCCA	92676	Db	93694	ATTAAGGTATCTGTTTGGAAATGGGATCATATTTAGGTGGGAAACATGATCTCTCG	93753
Qy	377	-----	377	Qy	494	PheIlePheCysIlePheGlyGlnPrometCysValLeuLeuTyrTyrHisAspLeuMet	513
Db	92677	TCTCTTTCAAAATTAATTTGCAAAATTCGAAAAACGAAAAAGCTAATCTCATACGAAT	92736	Db	93754	TTTATCTTCTGCAATTTTGGACAAACCGATGTGTGCTCTTATTACCAACGACCTGATG	93813
Qy	377	-----	377	Qy	514	AsnArgLysGlySerMetSer	520
Db	92737	TGATATTTTAGTTCTTAGAGTCGGTATATTTAGTTAGTTACTGACGCAAAATCTCT	92796	Db	93814	AAACGAAAAAGATCGATGTCA	93834
Qy	378	-----LeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLys	394	RESULT 13			
Db	92797	GTCCAAAGCTTAACATATTGGCAGAGCTTCTGCTCGGGGATCGTAATTCATCAAA	92856	AY084052	2090 bp	mRNA	linear, PLN 08-APR-2002
Qy	395	AspTyrPheAsnAlaLysSerValGlyAsp	404	LOCUS			
Db	92857	GATTTGTGGAAATGCAAAAGATGGGAGA-TGTGACCTATTCTTACTCAAAAGAAACTTA	92915	DEFINITION			
Qy	404	-----	404	complete cds.			
Db	92916	TGATTTTATATGTGTGCTGTTTGGTGCATCTAATACCAAAATTCATGATTCAC	92975	ACCESSION			
Qy	405	-----TyrTTPArgMetTyrPheMet	411	VERSION			
Db	92976	GTCTTCCTTTATACAGTACGTGAGAAATGTGAATATGTATGTCTCTTCTTAACATCA	93035	KEYWORDS			
Qy	411	-----	411	ORGANISM			
				Tropaeolum majus (nasturtium)			
				Tropaeolum majus			
				Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
				Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
				Rosids; eustosids II; Brassicales; Tropaeolaceae; Tropaeolum.			
				1 (bases 1 to 2090).			
				Characterization of a putative diacylglycerol acyltransferase mRNA			
				from Tropaeolum majus embryo			
				JOURNAL			
				Unpublished			
				2 (bases 1 to 2090)			

ASIDFERRAAANRRRLMDPHVVKHENSPPFIIDNPLSNPSTWCOFFRNAE  
LEKTDLDLRLPEHMCYFOTPRYQGLRLILWLCKLHEVYRQGMGLLAPLY  
VLHVDIMRLSEVRKSYEDYFTDRDPSLFRMERDITTFDEKFSYDNGGSGSHK  
NFNSLIDELDEVOGLVMTDYSYSESLGLVLSERFNEHDAVCFADMSGIRGCFAM  
ASFESYSPASGSHGLPLVLEACSAFRIALVADSSLSHSLVELGEVOYGLMLRY  
LFGEEFLIDOLLYWDEILILADNSARPDDESRNONRPIFPCDPRCTVIGMVSMLYL  
RSSLSLENTATCCLORLNFPENDILAKITOKAKLIALVALDITDMLASLNGVFDOS  
NFYARUKSCSTSPRPLIILAPESYWEKMWLVLAKEEENKICLERQTPPTOKKRM  
LNVKLERAVIDLSHNRKIGIGERKANSPPVQSLLEDSBQLNDCVHTVKNENIHQ  
ETEENIMEFHSADESIVSGSPSEESFVSLDPTSPVRCSTKIENDSVSAGSNLIP  
DEDEKSVISSEENSYSVSDPISPAIDSNKYKLDCTGSEMKDOOTSVNSLAYS  
PHRREXVPTQSDSDVSDKSVGTKEYKILHGIQVOPRKLKRLISEENSHRASQA  
TKTMDVKIKKNOICYSSESHQALSSDSSNLRKTLKNQOSMLKHAIELEYQOE  
PCLVPAGLIGLAKTNLEKQVATATLALKEKLSNLSME\*

gene  
/note="At2g19250"  
/note="synonym: F27F23.5"  
join(<17068, 17148, 18012, 18127, 18236, 18341, 18392, >18550)  
/gene="At2g19250"  
join(17068, 17148, 18012, 18127, 18236, 18341, 18392, 18550)  
/gene="At2g19250"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC16455.1"  
/db\_xref="GI:3135255"  
/translation="MNDENKNFVCVCRDSRREPIDEEQSYRCLILNIVERGIS  
FRNFCRLVAIYQPEGAIODEDPAVAIGAIKQLOEHCHEVRMYKETSNSHCNV  
KLVSILVLYKCSMTMWISLKWTLRPMVIMILKRRKETALVAMEKLAKMAY"  
complement(117459, 117479)  
/rpt\_family="(TAAAAA)n"  
/rpt\_family="AT\_rich"  
20539, 20576  
/rpt\_family="AT\_rich"  
21713, 24269  
/rpt\_family="(GA)n"  
/gene="At2g19260"  
/note="synonym: F27F23.6; predicted by genscan"  
join(<21713, 23329, 23377, 23431, 23522, 23625, 23778, 23868, 23909, 23941, 24091, 24215, 24233, >24269)  
/gene="At2g19260"  
join(21713, 23329, 23377, 23431, 23522, 23625, 23778, 23868, 23909, 23941, 24091, 24215, 24233, 24269)  
/gene="At2g19260"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="AAC16456.1"  
/db\_xref="GI:3135256"

## Alignment Scores:

Pred. No.: 4,22e-174 Length: 97495  
Score: 2225.00 Matches: 515  
Percent Similarity: 51.04% Conservative: 0  
Best Local Similarity: 51.04% Mismatches: 5  
Query Match: 80.30% Indels: 493  
DB: 8 Gaps: 15

US-09-623-514a-2 (1-520) x AC003058 (1-97495)

Oy 1 MetalalleuaspSerAlaGlyValThrValThrgluAsnGlyGlyGluPhe 20  
Db 90818 ATGGCGATTTTGGATTCTGCGCGCTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTTC 90877

Oy 21 ValAspleuaspArgLeuArgArgLysSerArgSerAspSerAsnGlyLeuLeu 40  
Db 90878 GTTCATCTTGTAGGCTTCGTCGACGGAATCGAGATCGATCTTCTTACAGCACTCTT 90937  
Oy 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60  
Db 90938 CTCTGTGCTCCGATTAATTAATCTCTCTCGATATGTTGAAGCTCCCGCGAGCTTAAG 90997  
Oy 61 AspArgIleAspSerValAlaAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
Db 90998 GATGGATGTATTCGCTGTTAAAGATAGCCTCAGGAAACAGCAATTTGCCGAGAT 91057  
Oy 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyAlaArgGlyGlyGlyGlyGlyAsn 100  
Db 91058 AATAACGGTGGTGGATTAATTAACGGTGGTGAAGAGCGCGGAGAAAGAGAAAC 91117  
Oy 101 AlaAspAlaThrPheThrValArgProSerAlaProAlaHisArgAlaArgLysSer 120  
Db 91118 GCCATGCTACGTTTACGATACGACCGCTGGTCCAGCTCATCGAGGCGAGAGAGAGT 91177  
Oy 121 ProLeuSerSerAspAlaIlePheLys----- 129  
Db 91178 CCACCTTAGCTCCGACGCAATCTTCAACAGGTTAAATCTCAGAAATTTGG 91237  
Oy 130 -----GlnSerHis 132  
Db 91238 TGTTCCTGTTGTTTATATGAAATGAGTTGCTGATGTTTGCATTCAGAGCCAT 91297  
Oy 133 AlaGlyLeuPheAsnLeuCysValValLeuIleAlaValAsnSerArgLeuIleLe 152  
Db 91298 GCCGATTAATCAACCTCTGTGTAGTAGTCTTATGCTGTAACAGTAGACTATCATC 91357  
Oy 153 GluAsnLeuMetLys----- 157  
Db 91358 GAAATCTTATGAAGGTTTGTCTACTTGTCTCTCTTTAGGAATGATGCTTGA 91417  
Oy 158 -----TyrGlyThrPleuIle 162  
Db 91418 AATTATCAGACAGAAATACCTTGTGTCATTCATTCATGATGATGATGAT 91477  
Oy 162 eArgThrAspPheThrPheSerSerArgSerLeuArgAspThrProLeuPheMet----- 180  
Db 91478 CAGACGATTTGTTGTTTATGTTCAAGATCGCTGGAGATGGCGGCTTTCAATGTCTG 91537  
Oy 180 ----- 180  
Db 91538 GTAAAGAGATGTTTTATTTCAGCAATGTTACATGTTATACGTATATGATAGT 91597  
Oy 181 -----CysCysIleSerLeuSerIlePhe 188  
Db 91598 TTAGTGATCAAGTCCCTCTTGAATCTCTCTCTGTCAGTATATCCCTTTCATCTT 91657  
Oy 188 eProLeuAlaIlePheThrValGlyLysLeuValLeuGlnLysTyrIleSerGluPro-- 207  
Db 91658 TCCTTTGGCTCTCTTTACGGTTGAGAAATGTTGTTCAAGAAATCATATCAGAACCTGT 91717  
Oy 207 ----- 207  
Db 91718 GAGTAATTAATCTTCACGCAATTAATTTATTAAGACAAATTTGATATCATG 91777  
Oy 208 -----ValValIlePheLeuHisIleIle 216  
Db 91778 AAGAACTTACAAGTCTCTTTGAATAATGCTCAAGGTTGTCATCTTCTCATATTAATA 91837  
Oy 216 IeThrMetThrGluValLeuLysProValIyValThrLeu----- 229  
Db 91838 TCACCAATCAGACAGGTTTGTATCCAGTTACGTCACCCATAGGTGATATGTTTCTG 91897  
Oy 229 ----- 229  
Db 91898 GTCTCAGTTTGTATCTGTTTAAAGTTAGTTGTCTGACCCGATATCTTGAATAATG 91957  
Oy 230 --ArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValT 249

Db 24171 AACGGAAGATGATGCA 24191

RESULT 12

AC003058

LOCUS

DEFINITION Arabidopsis thaliana chromosome 2 clone F27F23 map C1C06E08.

ACCESSION AC003058

VERSION AC003058.3

KEYWORDS GI:20197007

SOURCE HTG.

ORGANISM Arabidopsis thaliana (thale cress)

REFERENCE 1

AUTHORS Rounsley,S.D., Lin,X., Ketchum,K.A., Crosby,M.L., Brandon,R.C., Sykes,S.H., Kaul,S., Mason,T.M., Kerlavage,A.R., Adams,M.D., Somerville,C.R. and Venter,J.C.

JOURNAL Unpublished

REFERENCE 2

AUTHORS Lin,X.

TITLE Direct Submission

JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA

REFERENCE 3

AUTHORS Town,C.D. and Kaul,S.

TITLE Direct Submission

JOURNAL Submitted (27-FEB-2002) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA, cdtown@igf.org

COMMENT On Apr 18, 2002 this sequence version replaced gi:6598388.

FEATURES

source

1. 97495

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/cultivar="Columbia"

/db\_xref="taxon:3702"

/map="C1C06E08"

/chromosome="2"

/clone="F27F23"

1. 9642

/note="overlap with BAC clone T20K24 (AC002392:95887..105528)."

complement(21..3689)

/gene="At2g19210"

/note="synonym: F27F23.1"

complement(join<21..429,506..695,770..832,927..1053,1129..1340,1428..1594,1710..1790,1882..1953,2038..2109,2186..2324,2406..2884,2963..3509,3602..>3689))

/gene="At2g19210"

complement(join(21..429,506..695,770..832,927..1053,1129..1340,1428..1594,1710..1790,1882..1953,2038..2109,2186..2324,2406..2884,2963..3509,3602..>3689))

/gene="At2g19210"

/codon\_start=-1

/product="putative receptor-like protein kinase"

/protein\_id="AAC16451.1"

/db\_xref="GI:3135251"

/translation="MVHYNFLSLIIFACFAVVLVRAODSGEVSIDGIPEDSSYNDETDTIKYSDAFAVESCTIHSIDFQFSSLEKQPNRSPREGKNCYDKPQGGKGRYLIRTFMGNTDNDGKAPDFDITGFINWSDVTIDNATITVKEILIRLSH VHVCLVDKNGTPTPLSLLEIRLKSNTYEPDGLIFKRWDLGGLALVRKDDVFDRJWIPLRKGYTIFNASLITDSNNNGEQPARFVNMTATSPEDLSQDITFSEKDP TMYFYVMHAEVLEPNSNETREKYLINKEIKMSSESPRYLTDTLFGONPSVSGP LEERLOOTPRSTPLPIINATETRYRVNEFLOSPDODVDVIMRIKSYGKSMIGDP CAVKYPMKINDCSYVNDSPRIISVNLSSGIGETDAFNSUTLIIHIDLSNSIT GKTPTDFGNLHNTETLNECNKLSGALPVALLERSNKLILRLIDGPDLCVASCOI SDEKTKNVNITIPLVASVGLVLVLAIFLLYKKRHRGSGGVNAGDLDTTKRY KYSEVAVTNNEFRVNGKGFYVHGVLDNDQVAVKIISSSSAOGYKEPRAVEYELL RVHKNLTALIGCHEGKEMALIEFANGTLGDIYSGEKSYLSMERLOISLDAAG LEIYHNGCKPPIVORVDPANILINELQAKITADGLSRVALDGNNDPTVAGTIGYIDPEYHLTKQISEKSDIYSPGVILLEVSGOPVITARSNTAENHIHTRDVLNST GDIRGIVDPRLGGERPDAGSAMKITTEYAMACASSSKRPMTSHVVALKESSVRARAG

repeat\_region

gene

complement(5018..5047)

/rpt\_family="AT\_r1ch"

complement(5060..6543)

/gene="At2g19220"

/note="synonym: F27F23.2; predicted by genSCAN"

complement(join<5060..5590,5665..6171,6262..>6543))

/gene="At2g19220"

complement(join(5060..5590,5665..6171,6262..6543))

/gene="At2g19220"

/codon\_start=-1

/product="hypothetical protein"

/protein\_id="AAC16452.1"

/db\_xref="GI:3135252"

/translation="MTSKAAPEPHEFVDFLCVEOKMLGNQEMOHLIARFQMGVRFITDOLINHMDTMMKWKICRLVOCDDITKIDSLTNGKATDQEMAYILEVNEAGY RKNPPLFLELEITIFAGMINDGETSGSKMKQICERDENETGVIRIGLADYTHQ LKHFRTRTSMKHCETIASPIKMKWPNRKGATDEMDKYLIMINKRARVYKRII PHADKLATIEFGRIEPEKTYKTRRYKRVIDHSSPQLHDHQPSSVYVNTPEPVG SDRADKGVNPEPSTLIRSDSEDAVETVPEMMEKIPVNASVKKKEKTFEECYECCLDA IEEYKGGDLVMEALDLEFKRDYRYLFLMLQSKSLRANMLKLISN"

complement(5603..5623)

/rpt\_family="AT\_r1ch"

complement(9141..12813)

/gene="At2g19230"

/note="synonym: F27F23.3"

complement(join<9141..9558,9652..9841,9925..9987,10073..10199,10282..10478,10565..10722,10838..10918,11004..11075,11166..11237,11314..11452,11535..12019,12091..12637,12729..>12813))

/gene="At2g19230"

complement(join(9141..9558,9652..9841,9925..9987,10073..10199,10282..10478,10565..10722,10838..10918,11004..11075,11166..11237,11314..11452,11535..12019,12091..12637,12729..12813))

/gene="At2g19230"

/codon\_start=-1

/product="putative receptor-like protein kinase"

/protein\_id="AAC16453.1"

/db\_xref="GI:3135253"

/translation="MGNFNPFLVPSFASPVVLYVCAQDQSGEVSIDGIPEDSSYTDKTDIYIRMDAIVESGTHSIDSKQKKNLEDFQKVSFPEGKNCYDVPPOCK GFKYLIRTFMGNTDNDGKAPDFDITGFINWSDVTIDNATITVKEILIRLSH VHVCLVDKNGTPTPLSLLEIRLKSNTYEPDGLIFKRWDLGGLALVRKDDVFDRJWIPLRKGYTIFNASLITDSNNNGEQPARFVNMTATSPEDLSQDITFSEKDP TMYFYVMHAEVLEPNSNETREKYLINKEIKMSSESPRYLTDTLFGONPSVSGP LEERLOOTPRSTPLPIINATETRYRVNEFLOSPDODVDVIMRIKSYGKSMIGDP CAVKYPMKINDCSYVNDSPRIISVNLSSGIGETDAFNSUTLIIHIDLSNSIT GKTPTDFGNLHNTETLNECNKLSGALPVALLERSNKLILRLIDGPDLCVASCOI SDEKTKNVNITIPLVASVGLVLVLAIFLLYKKRHRGSGGVNAGDLDTTKRY KYSEVAVTNNEFRVNGKGFYVHGVLDNDQVAVKIISSSSAOGYKEPRAVEYELL RVHKNLTALIGCHEGKEMALIEFANGTLGDIYSGEKSYLSMERLOISLDAAG LEIYHNGCKPPIVORVDPANILINELQAKITADGLSRVALDGNNDPTVAGTIGYIDPEYHLTKQISEKSDIYSPGVILLEVSGOPVITARSNTAENHIHTRDVLNST GDIRGIVDPRLGGERPDAGSAMKITTEYAMACASSSKRPMTSHVVALKESSVRARAG

QY 188 eProleuAlaIaphethValGluLysLeuValLeuGlnLysTyrIleSerGluPro-- 207  
 |||||  
 Db 22015 TCCTTGGCTGCTTACGGTTGAGAAATTGGTACTTCAAAATATCATATCAGAACTGCT 22074  
 QY 207 ----- 207  
 Db 22075 GAGTAATTACTATTCTCCAGCCATTACTGTAATTTTATTGAAGACAAGCTTGTATCANG 22134  
 QY 208 -----ValValIlePheLeuHisIleIleI 216  
 |||||  
 Db 22135 AAGAACTTACAAGTCTGTTTGAAGAAAGCTCAAGTTGTCAATCTTCTTCATATATATA 22194  
 QY 216 IethmetHrGluValLeuTyrProValTyrValThrLeu----- 229  
 |||||  
 Db 22195 TCACCATGACAGAGTTTGTATCCAGTTACGTACCCCTAAGGTGATCTGTTTTCTG 22254  
 QY 229 ----- 229  
 Db 22255 GTCTCAGTTTGTGATCTGTTTTTAACTTTAGTTGTGACCCCGGTGATCTGAAAAATGG 22314  
 QY 230 --ArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuThrCysIleValT 249  
 |||||  
 Db 22315 ACAGGTGATCTGCTTTTATAGGTGTCACCTTGATGGTCCCTCAGCTTGATGTGT 22374  
 QY 249 rPLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAla 269  
 |||||  
 Db 22375 GGCTAAAGTTGGTTCTTATGCTCATACATGCTATGACATAGATCCCTTAGCCAAATGCAG 22434  
 QY 269 lAspPlySAla----- 272  
 |||||  
 Db 22435 CTGATTAAGCT-AAAAATACAAAAAGAGCGTATGTTAGTCACTTGCACTGCTGTACTG 22493  
 QY 273 -----AsnProGluValSerTyrTyrValSerL 282  
 |||||  
 Db 22494 TTTTAACAACAACCTGTATGAACCTTACGCCAATCTGAAAGCTCTCACTAGCTTAGCT 22553  
 QY 282 euLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyr----- 296  
 |||||  
 Db 22554 TCAAGAGCTTGGCATATTTTCATGTCGTCCTCCACATGTGTATCAGTAAGTCAAAAGT 22613  
 QY 296 ----- 296  
 Db 22614 GCATCAACCATCTTATCTTGCAGAGATTTCTGTGCTTAAACCTCGGATCTTTCCTTTC 22673  
 QY 297 --GlnProSerTyrProArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheA 316  
 |||||  
 Db 22674 CCCAGCCAAAGTATCCAGCTTCTGCATGTATACGGAAGGTTGGGTGCGTCATTTGG 22733  
 QY 316 lAlysLeuValIlePheThrGlyPheMetGlyPheIleIleGlu----- 330  
 |||||  
 Db 22734 CAAAACTGGTCATATTCACCGGATTCATGGATTATATATAGAAACAAAGTATTCACA 22793  
 QY 330 ----- 330  
 Db 22794 TCTTGCTTATTAAGTTTCTCTGGTGAATAATCATCATCCCTGCGTTGTACCACTTGACT 22853  
 QY 331 -----GlnTyrIleAsnProIleValAlaArgAsnSerLysHis 342  
 |||||  
 Db 22854 TCATGTTCTTTGTATCATTTTGGCAGATATATAAATCCTATGTGCGAAGACTCAAGCAT 22913  
 QY 343 ProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsn 362  
 |||||  
 Db 22914 CCTTTGAAAGGCGATCTTCTAATGCTATATGTAAGAGGTGTAAGCTTTCAGTTCCAAAT 22973  
 QY 363 LeuTyrValIlePheLysMetPheTyrCysPhePheHisLeuTyr----- 377  
 |||||  
 Db 22974 TTATATGTGTGGCTCTGCATGTCTTACTGCTCTTCCACCTTGGATGCTGTGATCCA 23033  
 QY 377 ----- 377  
 Db 23034 TCTCTTTCAAAATTAATTTCGAATTCGAAAAACCGAAAAAGCTAAATCTCATACGAAT 23093  
 QY 377 ----- 377

Db 23094 TGATATTTTAGTTTCTAGAGTCGGTGATGTAAATTTCAGTTACTGAAAGCAATCTCT 23153  
 QY 378 -----LeuAsnIleLeuAlaGluLeuLeuCysPheGlyValProGluPheTyrLys 394  
 |||||  
 Db 23154 GTCCAAAGGTTAAACATATTGGCAGAGCTTCTCTGCTGGGGATGTGAAATTCTACAAA 23213  
 QY 395 AspTyrPyrAsnAlaLysSerValGlyAsp----- 404  
 |||||  
 Db 23214 GATGTGTGAATGCAAAAAGTGTGGAGA-TGTGACCTATTTTACTCAAAAGAAACTTA 23272  
 QY 404 ----- 404  
 Db 23273 TGATTTTAAATGTGTGCTGTGTTTGGGTCAATCTAATCAACCAAAATCATATTCAGT 23332  
 QY 405 -----TyrTrpArgMetTyrAsnMet----- 411  
 |||||  
 Db 23333 GTCTTCTTTATCAGTACTGGAGAAATGTGAATGTGATGTCTCTTCTTAACATCA 23392  
 QY 411 ----- 411  
 Db 23393 CCTCTTTGTACACAAAAATAGAAAGAGCTAATTAAGATCTGTTTCTCTGACAG 23452  
 QY 412 ProValHisLysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIlePro 431  
 |||||  
 Db 23453 CCTGTCAATAAATGATGTGTGCATATATATCTCCGCTTGGCAGCAAGATACCA 23512  
 QY 432 Lys----- 432  
 Db 23513 AA-GGTGAGTGAGATATATACCGATATGCAATTTGTCGAATTTGTTCTGTGATATAAT 23571  
 QY 433 -----ThrLeuAlaIleIleIleAlaPheLeuValSer 443  
 |||||  
 Db 23572 TTAACTTCACACACTGTTGTTTTCAGACACCTGCATATATCTTCTCTAGCTCT 23631  
 QY 444 AlaValPheHisGlu----- 448  
 |||||  
 Db 23632 GCAGTCTTTCATGAGGATATACATATCTTCTACATTCCTGCTCTAGACGATGAACAC 23691  
 QY 448 ----- 448  
 Db 23692 ACCGTAGTGAAGAAATGCTAATATTCMAAGCATTTGTTTACTTAACGATCTGTGTTA 23751  
 QY 449 -----LeuCysIleAlaValProCysArgLeuPheLysLeuTyrAla 462  
 |||||  
 Db 23752 CAAATTCCTTTGACAGCTATGATCGCAGTTCTGTGCTGCTTCAAGCTATGGGCT 23811  
 QY 463 PheLeuGlyIleMetPheGln----- 469  
 |||||  
 Db 23812 TTTCTTGGATTAATGTTTCAGGTTAAAAAATTACTAACTGCTGCAGTCGATTTTACTA 23871  
 QY 470 -----ValProLeuValPheIle 475  
 |||||  
 Db 23872 AACTCAATCTCATATTCGTGACCAACCAATTTGTTGAGTAGTGCTTGGCTTTCATC 23931  
 QY 476 ThrAsnTyrLeuGlnGluArgPheGlySerThrVal----- 487  
 |||||  
 Db 23932 ACAAACTATCTACAGAAAGGTTGGCTCAACGGT-ATGCTCTCAAAACCCGAAATA 23990  
 QY 487 ----- 487  
 Db 23991 GAACGAATACTCTTCTTTCATAGCCTAGCCATTTAAATCGCAATGCGAAAATTATA 24050  
 QY 488 -----GlyAsnMetIlePheTyr 493  
 |||||  
 Db 24051 ATTAAGGTATCTGTTTGGATGGATGATATATATTAGCTGGGAAACATGATCTTCTGG 24110  
 QY 494 PheIlePheCysIlePheGlyGlnPrometCysValLeuLeuTyrTyrHisAspLeuMet 513  
 |||||  
 Db 24111 TTCACTTCTGCATTTTGGGACACGAGATGTGTGCTCTTATATACACAGACCTGATG 24170  
 QY 514 AsnArgLysGlySerMetSer 520  
 |||||



repeat\_region  
16808..16936  
/rpt\_family="Ar\_rich"  
complement(17023..17050)  
/rpt\_family="Ar\_rich"  
17139..17462  
/rpt\_family="Rr:ATREP5|ATREP5 An Ar-rich repetitive  
sequence - a consensus."  
complement(17361..17388)  
/rpt\_family="Ar\_rich"  
17595..17657  
/rpt\_family="Ar\_rich"  
18967..18992  
/rpt\_family="(TA)n"  
19827..19874  
/rpt\_family="(TA)n"  
20755..20797  
/rpt\_family="(TAAAAA)n"  
complement(20851..20873)  
/rpt\_family="Ar\_rich"  
21014..24475  
/gene="At2g19450"  
/note="synonym: F3P11.5; supported by cDNA:  
g\_11545079.gb,At054460.1,  
j01n121014..21564,21649..21729,21821..21894,21997..22072  
22170..22237,22319..22442,22523..22600,22679..22780,  
22881..23017,23162..23243,23348..23368,23453..23515,  
23599..23646,23770..23832,23914..23964,24090..24475)  
/gene="At2g19450"  
j01n121175..21564,21649..21729,21821..21894,21997..22072  
22170..22237,22319..22442,22523..22600,22679..22780,  
22881..23017,23162..23243,23348..23368,23453..23515,  
23599..23646,23770..23832,23914..23964,24090..24194)  
/codon\_start=1  
/product="diacylglycerol O-acyltransferase"  
/protein\_id="AAD1044.2"  
/db\_xref="GI:20197480"  
/translation="MALIDSAGVTVTYVTEGGGEFYVDLRLRRKRSRSDSSNGLLSGS  
DNNSDDVDGAPADVARDIDSVNDADQGTALADNNGGDNNGGGGEGRGND  
AFETVSPFAHRRARESPISDIAFKQSHAGLFMLCYVLIAVNSRLILENLMYGN  
LRTDFEFSRSLRDMPLMECCISLIFPLAFTYEKLVLOKYEISEPVYHITIM  
TEVLAIPVYVTELRDSAPLISGVTMLTCTIVWKLIVSAHSTSDISLANADKRAPEV  
SYVSIKSLIATFMAVPTLCQPSYPRSAICIRKGVAROPAKIVITFGWGTITEDYN  
PIVNRKAPLKGDLIYALERVLKLSVPNIYWLKMFYCFHLMWITAEILACFGDER  
YVDMNNAKSVGDGWRMNMPVHKMWVRHIIYECPLSKIPKLAIIALVAVAEFLC  
IADVPCLEFLFELAFGLIMEQVPLVFTYVLOERFSGTVGNMIFWFICIFGPMCVILY  
YHDLNMRKSGMS"  
25114..25134  
/rpt\_family="Ar\_rich"  
complement(25990..26015)  
/rpt\_family="Ar\_rich"  
26221..26844  
/gene="At2g19460"  
/note="synonym: F3P11.6; predicted by genefinder;  
supported by cDNA: g1\_14334633.gb,At034990.1,  
26221..26844  
/gene="At2g19460"  
26320..26670  
/gene="At2g19460"  
/codon\_start=1  
/product="expressed protein"  
/protein\_id="AAD1045.1"  
/db\_xref="GI:4191776"  
/translation="MEDYRSKSYGCGRISDLOQYSAHRRSDGPDGSGNQDRLSYIS  
TSTWTDPIKPEDONKRRKSSSSSSSGWGFVDPDQRRKRVYSYRAYTEGKLKGSFRK  
SEFKWIKDKCNKILN"  
complement(26742..26772)  
/rpt\_family="(CA)n"  
complement(28356..31273)  
/gene="At2g19470"  
/note="synonym: F3P11.7; supported by full length cDNA:  
Ceres.2312.  
complement(j01n<28356..28814,28901..28989,29085..29148,

[illegible]

```

DB      1207 CATGATATCTTCCCTGCTGGCGATCAAGATACCAAGATACCGCATTAATTCCT 1266
QY      440 PheLeuAlaSerAlaValPheHisGluLeuGlyIleAlaValProGlyArgGluPheLys 459
DB      1267 TTCTAGCTCTGCAGCTCTTCATGATGATGATGATGATGATGATGATGATGATGAT 1326
QY      460 LeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeu 479
DB      1327 CTATGGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1386
QY      480 GlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheGlyIlePhe 499
DB      1387 CAGAGAAAGTTGGCTCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1446
QY      500 GlyAlaProMetCysValLeuLeuTyrTrpHisAspLeuMetAsnArgGlyGlySerMet 519
DB      1447 GGACACACCGATGTGTGGGCTTCTTATACCATGACATGATGATGATGATGATGATG 1506
QY      520 Ser 520
DB      1507 TCC 1509

RESULT 11
AC005917 92822 bp DNA linear PLN 27-FEB-2002
LOCUS     Arabidopsis thaliana chromosome 2 clone F3p11 map C1C06E08,
ACCESSION AC005917
VERSION   AC005917.3 GI:20197478
KEYWORDS  HTG.
SOURCE    Arabidopsis thaliana (thale cress)
ORGANISM  Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosids; eudicots II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 92822)
Lin.X., Kaul, S., Shea, T. P., Fujii, C. Y., Shen, M., Vanaken, S. E.,
Barnstead, M. E., Mason, T. M., Bowman, C. L., Ronning, C. M.,
Beilto, M. -I., Carrera, A. J., Creasy, T. H., Buell, C. R., Town, C. D.,
Nierman, W. C., Fraser, C. M., and Venter, J. C.
Unpublished
2 (bases 1 to 92822)
Lin.X.
JOURNAL   Direct Submission
AUTHORS   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
REFERENCE Medical Center Dr., Rockville, MD 20850, USA
AUTHORS   Town, C.D. and Kaul, S.
TITLE     Direct Submission
JOURNAL   Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
COMMENT   Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
FEATURES  On Apr 18, 2002 this sequence version replaced gi:6598497.
          Source
            1..92822
              Location/Qualifiers
                /organism="Arabidopsis thaliana"
                /mol_type="genomic DNA"
                /cultiivar="Columbia"
                /db_xref="taxon:3702"
                /chromosome="2"
                /map="C1C06E08"
                /clone="F3p11"
                1..27852
                  misc_feature
                    (AC003058:65644..97495).
                    repeat_region
                      complement(2151..2171)
                      /rpt_family="AT-rich"
                    repeat_region
                      complement(5234..5268)
                      /rpt_family="AT-rich"
                    repeat_region
                      5488..5517
                      /rpt_family="AT-rich"
                      6440..8622
                      /rpt_family="AT-rich"
                      /gene="At2g19420"

```

```

mRNA
CDS
  /note="synonym: F3p11.2; predicted by genscan"
  join(<6440..6458,6961..7071,8318..>8622)
  /gene="At2g19420"
  join(6440..6458,6961..7071,8318..8622)
  /gene="At2g19420"
  /codon_start=1
  /product="hypothetical protein"
  /protein_id="AAD10141.1"
  /db_xref="GI:4191772"
  /translation="MHKVPKRTDSELYNGYLAIVNGYGLRFTYKLENDICVSI
  YNTGTRKSRKRRIRINVRPKIAPANKKPKRKEIDMPESDEVLNLSGDDG
  GYIPLPKRSMLNLDPSEFDEQTMHVSFLRVGKLMFG"
  complement(7164..7241)
  /rpt_family="(TA)n"
  complement(9348..9369)
  /rpt_family="AT-rich"
  9624..9757
  /rpt_family="AT-rich"
  /gene="At2g19430"
  /note="synonym: F3p11.3; contains a Trp Asp (WD 40) repeat
  signature"
  join(<9956..10281,10369..10445,10527..10572,10700..10760,
  10849..10902,10993..11073,11165..11204,11477..11577,
  11750..11860,12390..>12479)
  /gene="At2g19430"
  join(9956..10281,10369..10445,10527..10572,10700..10760,
  10849..10902,10993..11073,11165..11204,11477..11577,
  11750..11860,12390..12479)
  /gene="At2g19430"
  /codon_start=1
  /product="putative WD-40 repeat protein"
  /protein_id="AAD10142.1"
  /db_xref="GI:4191773"
  /translation="MYGDATWMNDESEILKRETRPTVPTANAPPARISNPD
  EYVASSDGTLPFHSNLVSOSASFGSKGDVAYAEPEYVRABEGPAIVKFGED
  EDALLSCGDGRVGRWKWREFAISDVSLLHLENLKRLLELIPQHGPGVGLSPMP
  EIMNASVDPQSGSVFTAGDSCAYCWDVESGRIKMTFEGHEDLHTVVSRSASQILP
  GSEEDATIMDCIKGCKYKVGISODKRSRLRVSSMALDGSSEMLVCGGKMLALNLP
  ASECVQFPIPAHVQDVDFDEKOVAVAGVGIVDVISQFSHLCITFRSSSL"
  complement(12903..14545)
  /gene="At2g19440"
  /note="synonym: F3p11.4"
  complement(join(<12903..13516,13584..14083,14223..>14545))
  /gene="At2g19440"
  complement(join(12903..13516,13584..14083,14223..14545))
  /gene="At2g19440"
  /product="putative beta-1,3-glucanase"
  /protein_id="AAD10143.1"
  /db_xref="GI:4191774"
  /translation="MNLAFVYGVGIMGIVAVDGLGVNMGTMATHKLPKRVOMTKD
  NINIKYKFDADETMTSLSSGLEVYAIITNDOLKVMGSDRAKDKWYRKRVYTNFN
  GCVNITFVAVGNREPFKSYNGSFNLNLPALQNLQNALNEGLSSVAVATPLNADYV
  DSPSSNPVPSAGRPADIIIGQMTQIVPLFGLGNNSAPITINIVPLSLGNDQFNLVAV
  FDGAKPVDNDNCIAYTNVDANFDVLVSLKAVGADMPIIVGEVMPLEGCKHANSQS
  AYRFYNGILPLRGNGPLRPITYIEVLFGLBDDAKSIAPGEERHMGIFKFDGQF
  KEPIDLSGQGNKLLIGAEVNTYQPKKCMNTAKDKDTKLAANIDVACTPSDCTAIG
  YGSSCNTLDANGNSYAFNNYFOYKNDQEDACITGATITTKNISQGCNFPYQIYA
  STASSFSSSLVLLIAGWFLSLGVAFY"
  15295..17666
  /rpt_family="Rf:ATREPS/ATREPS An AT-rich repetitive
  sequence - a consensus"
  15590..15728
  /rpt_family="AT-rich"
  15912..16149
  /rpt_family="AT-rich"
  complement(16233..16264)
  /rpt_family="AT-rich"
  16486..16577
  /rpt_family="AT-rich"
  complement(16668..16701)
  /rpt_family="AT-rich"

```

DEFINITION Brassica napus putative diacylglycerol acyltransferase (DGAT1)  
 mNNA, complete cds.  
 ACCESSION AF164434  
 VERSION AF164434.1 GI:5579407  
 KEYWORDS  
 SOURCE  
 ORGANISM Brassica napus (rape)  
 Brassica napus  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.  
 REFERENCE 1 (bases 1 to 1512)  
 AUTHORS Nykiforuk, C.L., Laroche, A. and Weselake, R.J.  
 TITLE Isolation and Characterization of a cDNA Encoding a Second Putative  
 Diacylglycerol Acyltransferase from a Microspore-derived Cell  
 Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.  
 AF164434). (PGR99-158)  
 JOURNAL Plant Physiol. 121 (3), 1053 (1999)  
 REFERENCE 2 (bases 1 to 1512)  
 AUTHORS Nykiforuk, C.L., Laroche, A. and Weselake, R.J.  
 TITLE Direct Submission  
 JOURNAL Submitted (30-JUN-1999) Chemistry and Biochemistry, University of  
 Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4,  
 Canada

FEATURES  
 source location/Qualifiers  
 1..1512  
 /organism="Brassica napus"  
 /mol\_type="mRNA"  
 /strain="cultivar Jet Neuf"  
 /db\_xref="taxon:3708"  
 /note="microspore-derived cell suspension culture"  
 1..1512  
 /gene="DGAT1"  
 1..1512  
 /gene="DGAT1"  
 /EC\_number="2.3.1.20"  
 /codon\_start=1  
 /product="putative diacylglycerol acyltransferase"  
 /protein\_id="A04536.1"  
 /db\_xref="GI:5579408"  
 /translation="MALISGCVAVPPTENGVALDLRLHRRKSSSDSSNGLSDTS  
 DVGAAAEEDRDVDAEEBAGDANLAGDARESDAGDVFTYRPSVPAHRRRE  
 SPLSDAIRKOSAGLFLNLCYVAVNSRLIETLMKGMILRDEPMSSTLRMP  
 LKPCCLSLVFLPAFTVERKMYLQKRISEPAIILVITMTVLIPVYTLKDSAF  
 LSGVTLMLFLCYLWKLVSIAHTSYDIRLANSADVDEPISYVSLKLYPMVAP  
 LCYQSPSPRSPIRKGMVAQOLAKLIVFLMGFIIIOVINPIVRNSKPLKODLLYA  
 IERVKLISVNLIVLMKMFCEPHMLNLIAELICGDEEFYKDMNNASVGDYRMM  
 NNPVHKMVRHYVEPCLRIKIPVPAIIIAFLVSAVHELCIIVPRLFNLMAFNGIM  
 FQVPIVEITNFIQERGSVMGNMIFGSACIFQPMCGILLYHIDLMNRGMS"  
 BASE COUNT 344 a 343 c 363 g 462 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1,39e-183 Length: 1512  
 Score: 2309.50 Matches: 442  
 Percent Similarity: 88.10% Conservative: 17  
 Best Local Similarity: 84.84% Mismatches: 43  
 Query Match: 83.35% Indels: 19  
 DB: 8 Gaps: 5

US-09-623-514a-2 (1-520) x AF164434 (1-1512)

QY 1 MetAlIleuAspSerAlaGlyValThrThr---ValThrgluAsnGlyGlyGlu 19  
 DB 1 ATGCGATTGTTGGATTCTGGAGCGTCGCGTACCGCCGAGGAGAAAGCGC----- 51  
 QY 20 PheValAspLeuAspArgLeuArgArgGlySerArgSerArgSerArgSerArg 39  
 DB 52 GTCCGAGATCTGCACAGGCTCCACCGTGAATTCAGATTCTTCCAAAGGAGCTC 111  
 QY 40 LeuLeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspVal 59  
 DB 112 CTC-----TCCGACTACTTCCCGCTCGGAGAGATGTGGAGCTGGCGGCGGAA 159

QY 60 ArgAspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGly 79  
 DB 160 AGGATCGGGTGTATTCCTCCGCGGAGAGAGGCTCAGGAAACAGGAATTTTAACT--- 216  
 QY 80 AspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyValArgGlyGlyGlyGlyArgGly 99  
 DB 217 -----GGCGAGATGCGCGAAACTAGGGAATCCCGGAGGC----- 252  
 QY 100 AsnAlaAspAlaThrPheThrTyArgProSerValProAlaHisArgArgAlaArgGlu 119  
 DB 253 -----GATTAAGGTTTACCTATTCACACCTCGCTTCACTATCGAGAGAGAGAG 306  
 QY 120 SerProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGly 139  
 DB 307 AGTCCTCTGACCTCCGACGCTATCTCAACAAAGCATCGAGATGTTCAACCTCTGT 366  
 QY 140 ValValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyGly 159  
 DB 367 GTAGTGTCTTCTTGTGTTACAGTACATCATCATCAAAACCTCATGAAGTATGTGT 426  
 QY 160 TrpLeuIleArgThrAspPheTrpPheSerArgSerLeuArgAspTrpProLeuPhe 179  
 DB 427 TGCTGATCAGACTGATTTTGTGTTAGTTCATCCTTACGAGACTGCGGCTTTTC 486  
 QY 180 MetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGlyLeuVal 199  
 DB 487 ATGTGTCTCTTTCACCTTTCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 546  
 QY 200 LeuGlnLysTyrrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThr 219  
 DB 547 CTTCAGAAATTCATATCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606  
 QY 220 GluValLeuTyrrProValTyrrValThrLeuArgCysAspSerAlaPheLeuSerGlyVal 239  
 DB 607 GAGGTCTGTATTCAGCTCATACGTCACACTAGAGTGATTCCTCTTCTGTAGGTGTC 666  
 QY 240 ThrLeuMetLeuLeuThrcysIleValTrpLeuLysLeuValSerTyrrAlaHisThrSer 259  
 DB 667 ACCTTATGCTGCTCATCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726  
 QY 260 TyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrrTyrr 279  
 DB 727 TACGACATTAAGAACCTGCGCAATTCAGCTGAATGATCATCTGCAAACTCTCTACTA 786  
 QY 280 ValSerLeuLysSerLeuAlaTyrrPheMetValAlaProThrLeuLysTyrrGlnProSer 299  
 DB 787 GTTACCTGGAAGCTTGGCGTATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 846  
 QY 300 TyrProArgSerAlaCysIleArgLysGlyTyrrValAlaArgGlnPheAlaLysLeuVal 319  
 DB 847 TATCCACGTTCTCCATGATATCCGAAAGGTTGGTGGCTGCTCAACTGCAAAACGTGTC 906  
 QY 320 IlePheThrGlyPheMetGlyPheIleIleGluGlnTyrrIleAsnProIleValAlaGln 339  
 DB 907 ATATTACCTGACTCATAGGATTTATTAATGACCAATATTAATCTATTTGTAGAAC 966  
 QY 340 SerLysHisProLeuLysGlyAspLeuLeuTyrrAlaIleGluArgValLeuLysLeuSer 359  
 DB 967 TCAAAACATCTCTGAAAGGAGACCTTATATGCTATTAAGAAAGAGCTTTGAAGCTTTCA 1026  
 QY 360 ValProAsnLeuTyrrValTrpLeuLysMetPheTyrrCysPhePheHisLeuTrpLeuAsn 379  
 DB 1027 GTTCCAAATCTATATATGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1086  
 QY 380 IleLeuAlaGlnLeuLeuLysPheGlyAspArgGlnPheTyrrLysAspTrpTrpAsnAla 399  
 DB 1087 ATATTGGCAGAGCTCTCTGCTTGGGAGACGTAATTCACAAATTTGGTGAATGCA 1146  
 QY 400 LysSerValGlyAspTyrrTrpArgMetTrpAsnMetProValHisLysTrpMetValArg 419  
 DB 1147 AAAAGCGTGGAGATTTATGGAATGTGGAATATGCTGCTGCTGCTGCTGCTGCTGCTGCT 1206  
 QY 420 HisIleTyrrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAla 439



```

/translation="MAILDSAGVTVTENGGEFVDLRLRRKSSDSSNGLLSSGS
DNNSPSDDVADPADVRIDSVNDADAGANLANAGNNNGRGGERGNAD
AFETVSPVAHRRARSPSLSDAIFKOSVNLVYLIVNSLLIENMKMG
LIRTFDESSRSRLRDMPLFMCISLSTFPLAFTVEKIVLOXISEPVYELIITM
TEVLYPVYVTLRCDSDAFISGVITMLITCTIVKIVSAHNSYDRLSANAADANPEV
SYVYSKSLAFVAVAPTLCTQPSYPSACIRKGVAROFKLVITFGKFTIEQITN
PIYRNSKHPLGDLVLAIEVRLKSVNPLVLMCTYCFEHLNLIETLCLGDEEF
YKDMWNAKSVGDYRMMNMPVHKWVRIYFPLRSKIPKTLAIIIAFLVSAVHEIC
IAVPCRLFLKMLAFGLIMFOVPLVFITNYLQERGVSNMIFWIFICIFGPMCVILLY
YHDLNMRKGSMS"
BASE COUNT      499 a      413 c      447 g      646 t
ORIGIN
Alignment Scores:
Pred. No.:      5,28e-222      Length:      2005
Score:          2771.00      Matches:      520
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             8      Gaps:      0
US-09-623-514a-2 (1-520) x AY054480 (1-2005)
QY      1 MetAlAlIeLeuSpSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20
DB      162 ATGGCGATTTTGAATCTGCTGGCGTTACTACGTCGAGAGAACGGTGGCGGAGATTC 221
QY      21 ValAspLeuSpArGleuArGArGArGlySerArGSerArSpSerSerArSnGlyLeu 40
DB      222 GTGATCTTGAATAGCGCTCTCGACGGAATCGAGATCGGATCTTCTTAACGACTTCT 281
QY      41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
DB      282 CTCCTCGTCCCATATATATCTCCTTCGATGATGTTGAGACTCCCGCGGAGTTAGG 341
QY      61 AspArGleuSpSerValAlaAsnAspAlaGlyThrAlaAsnLeuAlaGlyAsp 80
DB      342 GATCGGATGATTCCTGCTGTTAACGATGACGCTGAGGAAACGCCAATTTGGCCGGAGAT 401
QY      81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 100
DB      402 AATTAAGCGGTGGTGGCATATTAACGGTGGTGGAAAGCGCGCGAGAGAGAGAGAGAAC 461
QY      101 AlaAspAlaThrPheThrTyraGProSerValProAlaHisArgAlaArgGluSer 120
DB      462 GCCGATGCTACGTATCGATCGACCGCTCCGATCATCGGAGCGGAGAGAGAGAT 521
QY      121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
DB      522 CCACCTTAGCTCCGACGCAATCTTCNAACAGACGATGCCGGAATTAATCAACCTCTGTGTA 581
QY      141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyGlyTrp 160
DB      582 GTAGTCTTATGCTGTGAACAGTACAGCATCATCGAANAATCTTATGAAGTATGTTGG 641
QY      161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
DB      642 TTATATAGAACGGAATTTCTGTGTAGTTCAGATCGCTGCAATTTGGCCGCTTTTCATG 701
QY      181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
DB      702 TGTGTATATCCCTTTCGATCTTCTTGTGGCTGCTTTCGCTTGAAGAAATGGTACTT 761
QY      201 GlnLysTyrlleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
DB      762 CAGAAATACATATACAGAACCTGTGTCATCTTCTTCATATATATACACATGACAGAG 821
QY      221 ValLeuTyrrProValTyrrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
DB      822 GTTTTGTATCCAGTTTACGTCACCCCTAAGGTGATTTCTCTTTTATATAGGTCACAT 881
QY      241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrrAlaHisThrSerTyrr 260

```

```

DB      882 TTGATGCTCCTCACTTCGATTTGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGTAT 941
QY      261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrrVal 280
DB      942 GACATATAGATCCCTAGCAATTCGACGATGATGAAGCCCATCTGAAGTCTCTCACTAGCT 1001
QY      281 SerLeuLysSerLeuAlaTyrrPheMetValAlaProThrLeuGlyTyrrGlnProSerTyrr 300
DB      1002 AGCTTGAAGAGCTTGGCATTTTCATGCTGCTGCCACATGATGTTATACGCCAAGTTAT 1061
QY      301 ProArgSerAlaCysIleArgLysGlyTyrrValAlaArgGlnPheAlaLysLeuValIle 320
DB      1062 CCACGTTCTCAATATACGTAAGGAGGTGGGTGGCTCTCAATTTGCAAAAGTGCATATA 1121
QY      321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrrIleAsnProIleValArgAsnSer 340
DB      1122 TTCACCGGATTCATGGGATTTATATATGAACCAATATATATCTTATTTGCAGAACTCA 1181
QY      341 LysHisProLeuLysGlyAspLeuLeuTyrrAlaIleGluArgValLeuLysLeuSerVal 360
DB      1182 AACCATCCTTTGAAGGCGATCTTCTATATGCTATGTAAGAGAGTTGAAGCTTTCAGTT 1241
QY      361 ProAsnLeuTyrrValTrpLeuGlySerMetPheTyrrCysPhePheHisLeuTrpLeuAsnIle 380
DB      1242 CCAAAATTAATATGTGGCTCTGCATATGTTCTACTGCTTCTCCACCTTGGTTAAACATA 1301
QY      381 LeuAlaGluLeuLeuGlyCysPheGlyAspArgGluPheTyrrLysAspTrpTrpAsnAlaLys 400
DB      1302 TTGGCAAGCTTCTCTCTCTCTCGGGATCGTCAATTCACAAAGTTGTCGAATGCCAAA 1361
QY      401 SerValGlyAspTyrrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
DB      1362 AGTGTGGAGATTAATACGGAAGAAATGTGGAATATGCTCTTCATTAATGATGTTGCACAT 1421
QY      421 IleTyrrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe 440
DB      1422 ATATACTCCCGGTGCTGGCCAGCAAGATACCAAAAGCACTCCCATTAATCATGCTTTC 1481
QY      441 LeuValSerAlaValPheHisGluLeuGlyCysIleAlaValProCysArgLeuPheLysLeu 460
DB      1482 CTAGTCTCTGACATCTTTCATGAGCTATGATGATGATGATGATGATGATGATGATGATGAT 1541
QY      461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrrLeuGln 480
DB      1542 TGGGCTTTTCTGTGGATATGATTTACAGTCCCTTGTGCTTCATCAAACTATCTACAG 1601
QY      481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
DB      1602 GAAAGGTTTGGCTCAACGGTGGGGAACATATATCTTCTGTTCAATCTTCTGATTTTCGA 1661
QY      501 GlnProMetCysValLeuLeuTyrrTyrrHisAspLeuMetAsnArgLysGlySerMetSer 520
DB      1662 CAACCGATGTGTGCTCTTATTAACACAGCTGATGAACCGAAAGAGATGAGTCTCA 1721
RESULT 9
AF251794      1537 bp      mRNA      linear      PLN 16-Apr-2000
LOCUS      AF251794
DEFINITION      Brassica napus putative diacylglycerol acyltransferase mRNA,
complete cds.
ACCESSION      AF251794
VERSION      AF251794
KEYWORDS      GI:7576940
SOURCE      Brassica napus (rape)
ORGANISM      Brassica napus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE      1 (bases 1 to 1537)
AUTHORS      Brown,A.P., Schiefer,T.P. and Slabas,A.R.
TITLE      Characterization of a putative diacylglycerol acyltransferase cDNA
from Brassica napus embryo
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 1537)

```



```

Db      717 TTGATCAGAACGATTTTCGCTTACGTTCAGATTCGCGAGATTCGCCGCTTTTCAG 776
QY      181 CysCys11SerLeuSer11LeuPheProLeuAla1AlaPheThrVal1GluLeuValLeu 200
      777 TGTGTATATCCCTTTCGATCTTTCCTTGGCTGGCTTTCGCTTGAAGATTCGACT 836
QY      201 GlnLysTrl1SerGluProVal11LeuPheLys11Leu11Leu11Leu11Leu11Leu 220
      837 CAGAAATACATATCAGAACCTGTTCATCTTCTTCATATATATATACACATGACAGAG 896
QY      221 ValLeuTrpProVal1TrpVal1ThrLeuArgCysAspSer11LeuLeuSerGlyVal1Thr 240
      897 GTTGTGATTCACATTTACGTCACCCATGAGGTGATTCGCTTTTATCAGGTGCTACT 956
QY      241 LeuMetLeuThrCys11LeuAla1TrpLeuLysLeuValSer1TrpAla1HisThrSer1Trp 260
      957 TTGATGCTCTCTACCTTCATCTGTGTGCTTAAAGTTGGTTCTTATGCTCTACTACTAT 1016
QY      261 Asp11LeuArgSerLeuAlaAsnAla1AlaAspLysAlaAsnProGluValSer1TrpTrpVal 280
      1017 GACATATAGATCCCTACCAATGACGATGATAGGCCAATCTCGAAGCTCTCTACTAGCTT 1076
QY      281 SerLeuLysSerLeuAla1TrpMetVal1AlaProThrLeuLysTrpGlnProSer1Trp 300
      1077 AGCTTGAAGAGCTTGGCATATTTTCATGTGCTGCTCCACATGTTGATATCAGCAAGTTAT 1136
QY      301 ProArgSerAlaCys11LeuArgLysGlyTrpVal1AlaArgGlnPheAlaLysLeuVal11Le 320
      1137 CCAGCTTTCGATGATATGAGGAAGGTGGTGGCTGCTGCTCAATTCGCAAAACGCTGCATA 1196
QY      321 PheThrGlyPheMetGlyPhe11Leu11Leu11Leu11Leu11Leu11Leu11Leu11Leu 340
      1197 TTCACCGGATTCATGGGATTTATATAGAACATATATATATATATATATATATATATAT 1256
QY      341 LysHis1ProLeuLysGlyAspLeuLeuTrpVal11Leu11Leu11Leu11Leu11Leu11Le 360
      1257 AAGCATCTTTGAAAGCCGATCTTCTATATGCTATGTAAGAGGTGTAAGGCTTTCAGT 1316
QY      361 ProAsnLeuTrpVal1TrpLeuLysMetPheTrpCysPhePheHis1LeuTrpLeuAsn11Le 380
      1317 CCAAATTTATATGTGGCTCTGCTCATGCTCTACTGCTTCTTCACCTTGGTAAACATA 1376
QY      381 LeuAlaGluLeuLeuLysPheGlyAspArgGluPheTrpLysAspTrpPheAsnAlaLys 400
      1377 TTGGCAGAGCTTCTGCTTGGGATCGGATTCATACAAATTCGTGGAATGCAAAA 1436
QY      401 SerVal1GlyAspTrpTrpArgMetTrpAsnMetProVal1HisLysTrpMetVal1ArgHis 420
      1437 AGCTGGGAGATTCAGTGAAGATGCGAATATGCTGCTTCATATATGATGTTGCGCAT 1496
QY      421 IleTrpPheProCysLeuLysSerLys11LeuProLysThrLeuAla11Leu11Leu11Le 440
      1497 ATTTACTTCCCGCTTGGCGACCAAGATATACCAAAACACTGCGCATTTATCATTCGTTTC 1556
QY      441 LeuVal1SerAlaVal1PheHisGluLeuCys11LeuVal1ProCysArgLeuPheLysLeu 460
      1557 CTAGTCTGCACTCTTTCATGAGTATGATGATGCGATTCGCTTCTCTTCAAGCTA 1616
QY      461 TrpAlaPheLeuLys11LeuMetPheGlnVal1ProLeuVal1Phe11LeuHisTrpLysGln 480
      1617 TGGGCTTTTCTTGGGATTTATGTTTTCAGGTGCCCTTTGGTCTTCATCAACAACTATCTAC 1676
QY      481 GluArgPheGlySerThrVal1GlyAsnMet11LeuPheTrpPhe11LeuPheCys11LeuPhe 500
      1677 GAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGCTTCATCTTCGCAATTTTCGGA 1736
QY      501 GlnProMetCysVal1LeuLeuTrpTrpHisAspLeuMetAsnArgLysGlySerMetSer 520
      1737 CAACCGATGTGTGTCTCTTATATACCAAGACCTGATGACGGAAGAGATGATGTA 1796

```

```

RESULT 7
ATH131831
LOCUS   ATH131831      1988 bp      mRNA      linear      PLN 10-JUN-1999

```

```

DEFINITION Arabidopsis thaliana mRNA for diacylglycerol O-acyltransferase.
ACCESSION AJ131831
VERSION AJ131831.1 GI:5050912
KEYWORDS diacylglycerol O-acyltransferase.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
REFERENCE 1
AUTHORS Hobbs,D.H., Lu,C. and Hills,M.J.
TITLE Cloning of a cDNA encoding diacylglycerol acyltransferase from
JOURNAL Arabidopsis thaliana and its functional expression
MEDLINE 99313150
PUBMED 10386579
REFERENCE 2 (bases 1 to 1988)
AUTHORS Hills,M.J.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1998) Hills M.J., Department of Brassica and
Oilseeds Research, John Innes Centre, Colney Lane, Norwich, NR4
7UH, U.K.
FEATURES
Source location/Qualifiers
1..1988
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/sub_species="Columbia"
/db_xref="taxon:3702"
231..1793
/EC_number="2.3.1.20"
/codon_start=1
/product="diacylglycerol O-acyltransferase"
/protein_id="CA44774.1"
/db_xref="GI:5050913"
/translacion="MAIIDSAGTYYTENGCGEPVDLRLRRKSRSDSSNGILLSS
DNNSPDVGAPEADVRIDSVNDQAGTANLANNNGGDDNGGKGBGSGNAD
AETTPSPVAHRRARESPSSDAIRKQSHAGFNLGVVLAIVNSRLIIEINLKGW
LRTDWFSSRSRLDWPLEWCCISLIFPLAATVVKLVOKYISPVVIFPHIITM
TEVLVYVYTLRCDSAFLSGVTLMLTICVLTLYVATSYDIRSLANADANAEV
SYVSLSLAYEVAPTLCYOPSPRACIRKGVARORAKLYIFGFMGFTIEOYN
PIYRNSKHPLKGLDLVAIRVYKLVSPIVYVWCMGCFPHMLNLALFLCGDEF
YKDMANAKSVGDVWRMNNVPHNKNRHHYFPELRKIKRTLAITIAFLVAVFELC
IAPCFLEKLMAFLEGTMFQVPLFTITNYLQERFGSTVGNNIFWIFCIGFQPCVILY
YHDLNMRKSSMS"
BASE COUNT 478 a 423 c 436 g 651 t
ORIGIN
Alignment Scores:
Pred. No.: 5,22e-222 Length: 1988
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-09-623-514a-2 (1-520) x ATH131831 (1-1988)
QY      1 MetAla11LeuAspSerAlaGlyVal1ThrVal1ThrGluAsnGlyGlyGluPhe 20
      231 ATGGCATTTTGGATTTCTGCTGCGTACTACGCTACGAGACGAGCGGAGAGTTC 290
Db      291 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeu 40
      291 GTGCATCTTGATVAGGCTTGTGCGACGGAATGAGATTCGATCTTAACGACTTCT 350
QY      41 LeuSerGlySerAspAsnAsnSerProSerAspVal1GlyAlaProAlaAspValArg 60
      351 CTTCTCGTTCGATATATATCTCTTGGAGATGTGGAGCTCCCGCCGACGTTAGG 410
QY      61 AspArg11LeuAspSerVal1AlaAsnAspAspAlaGlnGly1ThrAlaAsnLeuAlaGlyAsp 80
      411 GATCGATTTGATTCCTGTTTACAGTACGCTCAGGGAACACGCAATTTGGCGGAGAT 470

```



QY	141	ValValIleuIleAlaValAsnSerArgIleuIleIleGluAsnIleuMetLysTyrGlyTyr	160
Db	657	GTAACTTCTTAATGGTGTAAACAGTAGACTCATCATCGAAATCTTAATGAATATAGTTGG	716
QY	161	LeuIleArgIleAspPheTyrPhePheSerSerArgSerIleuArgAspTyrProIleuPheMet	180
Db	717	TTGATTCAGAAACGAGATTTCTGGTTTACGTTCAAGATCGCTCCGAGATTTGGCCCTTTTCAAG	776
QY	181	CysCysYllSerIleSerIlePheProIleuAlaAlaPheThrValGluLysLeuValIleu	200
Db	777	TGTTGTATATCCCTTTGGATCTTTCCTTTGGCTCCCTTACGGTTGAGAAATGGTACTT	836
QY	201	GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu	220
Db	837	CAGAAATATCATATCAGAAACCTGGTTGTCACTTTCTTCATATATATATACCATGACAGAG	896
QY	221	ValLeuTyrProValTyrValTThrLeuArgCysAspSerAlaPheLeuSerGlyValTThr	240
Db	897	GTTTTGTATCCAGATTATACGTCACCCCTAAAGGATGATTCGCTTTTATATCAGGGTCACT	956
QY	241	LeuMetLeuLeuThrCysIleValTTrpLeuLysLeuValSerTyrAlaHisThrSerTyr	260
Db	957	TTGATATGCCCTCACCTTGCAATGTGTGGCTAAAGTTGGTTCTTATATGCCATACATACGAT	1016
QY	261	AspIleArgSerIleuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyrVal	280
Db	1017	GACATTAAGATCCCTACGCAATGACAGCTGATAGGCCAATCCTGAAGTCTCTACTACGTT	1076
QY	281	SerLeuLysSerIleuAlaTyrPheMetValAlaProThrIleuCysTyrGlnProSerTyr	300
Db	1077	AGCTTTGAAGACCTTGGCATATATTCATGTGCTCCCTCCACATTTGTCTTACACCCAAAGTTAT	1136
QY	301	ProArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIle	320
Db	1137	CCAGCTTCTGCATGTATATACGGAAGCGTTGGTGGCTGTCATTTGCCAAACCTGCTATA	1196
QY	321	PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer	340
Db	1197	TTCCACCGGATTCATAGGAGATTTATATAGAACAAATATATTAATCCTATTTGTACAGAACATCA	1256
QY	341	LysHisIleProLeuLysGlyAspIleuIleTyrValIleGluArgValIleuLysLeuSerVal	360
Db	1257	AAGCATCTTTGAAAGCGCATCTTCTATATGCTATGTGAAGAGGTTGAAACCTTTCAGTT	1316
QY	361	ProAsnLeuTyrValTTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle	380
Db	1317	CCAAATTTATATGTGTGGCTCTGCATGTCTACATGCTTCTCCACCTTTGGTTAAACATA	1376
QY	381	LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAlaLys	400
Db	1377	TTGGCAGAGCTTCTCTGCTCTGGGGAGTCGTAAATTCACAAAGATGGTGAAGCAAA	1436
QY	401	SerValAlaAspTyrTTrpArgMetSerTrpAsnMetProValHisLysTrpMetValArgHis	420
Db	1437	AGTGTGGAGATTACTGAGAAATGTGGAAATGTGCTGTCTTAATATGATGCTTCGACAT	1496
QY	421	IleIleTyrPheProCysLeuArgSerLysIleProLysThrIleuAlaIleIleAlaPhe	440
Db	1497	ATATATCTTCCCGTCTTGGCAGCAACATATACCAAGACACCTGCCATTAATCATTTGCTTTC	1556
QY	441	LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu	460
Db	1557	CTAGTCTCTGACAGCTTTCATGACGTATGCAATGCGAGTTCTTGTCTCTTTAAGCTA	1616
QY	461	TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln	480
Db	1617	TGGCTTTTCTTCTGGAGATTAATGTTTCAGGTGGCTTTGGTTCATACACAAATATCTACAG	1676
QY	481	GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly	500
Db	1677	GAAGAGTTTGGCTCAACGGTGGGAAACATGATCTTCTGGTTCATCTTCGATTTTGGGA	1736

QY	Db	RESULT 6	AX090349	AX090349	1942 bp	DNA	linear	PAT 21-MAR-2001
			Sequence 42 from Patent WO01161308.					
			AX090349					
			AX090349					
			AX090349.1	GI:13444210				
			Arabidopsis thaliana (thale cress)					
			Arabidopsis thaliana					
			Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
			Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
			rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.					
			1					
			Lassner, M. and van Eenennaam, A.					
			Plant sterol acyltransferases					
			Patent: WO 0116308-A 42 08-MAR-2001;					
			MONSANTO COMPANY (US)					
			Location/Qualifiers					
			1..1942					
			/organism="Arabidopsis thaliana"					
			/mol_type="genomic DNA"					
			/db_xref="taxon:3702"					
			BASE COUNT	461 a	421 c	425 g	635 t	
			ORIGIN					
			Alignment Scores:					
			Pred. No.:	5.06e-222	Length:	1942		
			Score:	2771.00	Matches:	520		
			Percent Similarity:	100.00%	Conservative:	0		
			Best Local Similarity:	100.00%	Mismatches:	0		
			Query Match:	100.00%	Indels:	0		
			DB:	6	Gaps:	0		
			US-09-623-514A-2 (1-520) x AX090349 (1-1942)					
QY	1	1	MetAlaIleuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGlyUphe	20				
Db	237	ATGGCGATTTTGGATTTCTGCTGGCTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTC	296					
QY	21	ValAspIeuAspAlaGlyLeuArgArgArgLysSerArgSerSerSerSerAsnGlyLeuLeu	40					
Db	297	GTCGATCTTGATAGGCTTCGTCGACGGAGAAACGAGATCGGATTTCTTACGGGACTCTT	356					
QY	41	LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg	60					
Db	357	CTCTCTGGTCCGATTAATTAATTCCTTCCTCGATGATGTGGAGCTCCCGCGAGCTTAGG	416					
QY	61	AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp	80					
Db	417	GATCGGATTTGATTCCTGTTGTAACGAGACCCCTCAGGGACAGCCAAATTTGGCCGAGAT	476					
QY	81	AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyValArgGlyGlyGlyValArgGlyAsn	100					
Db	477	AATTAACGTGTGTGGCGGATTAATTAACGGTGTGTGTGAAGAGCGCGGAGAGGAGAGAAC	536					
QY	101	AlaAspAlaThrPheThrThrTyrArgProSerValProAlaHisArgArgAlaArgGluSer	120					
Db	537	GCCGATCTCTAGCTTTACGTATCGACCGCGCGTTCACGTCATTCGAGAGCGGAGAGAGT	596					
QY	121	ProIeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal	140					
Db	597	CCACTTGTAGCTCCGACGGAATCTTAAACAGAGCATCCCGGATTAATTCACCTCTGTGA	656					
QY	141	ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrP	160					
Db	657	GTAAGTTCTTATGCTGTAAACAGTAGACATCAATCCGAAATCTTATGAAGTAGGTGG	716					
QY	161	LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTyrProLeuPheMet	180					



QY 101 AlaaspaIatPhPheThyTyrArpProSerValProAlaHisArgAlaArgIuSer 120  
 |||||  
 Db 537 GCCGAGTCTACCTTTACGTATGACACCGCTGCCGTCCAGTCATCGAGCGAGAGAGACT 596  
 QY 121 ProLeuSerSerAspAlaIlePheIysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
 |||||  
 Db 597 CCACCTAGCTCCGACCAATCTTCAACAGACGCGGATTAATCAACCTCTGTGA 656  
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetCysTyrGlyTyr 160  
 |||||  
 Db 657 GTAGTCTTATGCTGTATAACAGTACATCATCATGAAATCTTATGAAGTATGTTGG 716  
 QY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180  
 |||||  
 Db 717 TTGATTCAGAACGATTTCTGTGTGTAGTCAAGATCCCTCGAGATTGGCCGCTTTTCAG 776  
 QY 181 CysCysIleSerLeuSerIlePheProLeuAlaIlePheThrValGluIysLeuValLeu 200  
 |||||  
 Db 777 TGTGTATATCCCTTTCGATCTTCTTGGCTGCCCTTACGGTGAATAATGGTACTT 836  
 QY 201 GlnIysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220  
 |||||  
 Db 837 CAGAAATACATATACAACTGTTGTCATCTTCTCATATATATATACACATGACAGAG 896  
 QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
 |||||  
 Db 897 GTTTGTATCCAGTTTACCTCAACCCCAAGCTTGATTCGCTTTTATCAGGTGCACAT 956  
 QY 241 LeuMetLeuLeuThrCysIleValIlePheLeuValSerTyrAlaHisTrpSerTyr 260  
 |||||  
 Db 957 TTGATCTCTCTACCTTTCATTTGTGTGCTAAGTTGTTCTTATCTCATCTACTACTCT 1016  
 QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspIysAlaAspProGluValSerTyrTyrVal 280  
 |||||  
 Db 1017 GACATTAAGATCCCTACCAATGACGTGATTAAGGCCAATCCGAACTCCACTACTACT 1076  
 QY 281 SerLeuIysSerLeuAlaIlePheMetValAlaProThrLeuCysTyrGlnProSerTyr 300  
 |||||  
 Db 1077 ACCTTGAAAGCTTGGCATATTTCAATGCGCTCCACCATTTGCTATACGCCAAGCTTAT 1136  
 QY 301 ProArgSerAlaCysIleArgIysGlyTyrValAlaArgGlnPheAlaIysLeuValIle 320  
 |||||  
 Db 1137 CCACGTTTGCATGTATACGGAAGGTTGGGTGGCTGCGCATTTTCCAAACATGGGTCA 1196  
 QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAspProIleValArgAsnSer 340  
 |||||  
 Db 1197 TTCACCGGATTCATGGGATTTAATAAGAACATATATAATCCCTTTCACGAGACTCA 1256  
 QY 341 LysHisProLeuIysGlyAspLeuLeuTyrAlaIleGluArgValLeuIysLeuSerVal 360  
 |||||  
 Db 1257 AAGCATCCCTTTGAAAGCGCATCTTCTATGCTATGCTAATGAAAGAGTCTTGAAGCTTTCAGTT 1316  
 QY 361 ProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380  
 |||||  
 Db 1317 CCAAATTTATATGTGTGGCTGTGCATGTCTACTGTCTTCTCCACTTGGTTAAACATA 1376  
 QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrIysAspTrpPheAsnAlaIys 400  
 |||||  
 Db 1377 TTGGCAGACTCTCTGCTTGGGGATTCGTGAATTTACAAAGATTGGCGAATGCAAAA 1436  
 QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisIysTrpMetValArgHis 420  
 |||||  
 Db 1437 AATGTGGAGATTTACGAGAAATGTGGATATGCCCTTCATAAAGATGGTTCACAT 1496  
 QY 421 IleTyrPheProCysLeuArgSerIysIleProIysThrLeuAlaIleIleIleAlaPhe 440  
 |||||  
 Db 1497 AATATCTCCCTGCTTGGCGAGCAAGATACCAAGACATCGCCATTAATCATTTGCTTTC 1556  
 QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheIysLeu 460  
 |||||  
 Db 1557 CTAGCTCTGCAAGTCTTTCATGAGCTATGCATCGCAGTTCCTGTGCTCTCAAGCTTA 1616  
 QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480

Db 1617 TGGGCTTTCTTGGGATGTATGTTTCAGTGCCCTTGTGCTCTATCAACAATCTACTACAG 1676  
 QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500  
 |||||  
 Db 1677 GAAAGTTTGGCTCAACGGGTGGGGAACATGATCTTGTGTCTATCTTGTGCAATTTTCGA 1736  
 QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgIysGlySerMetSer 520  
 |||||  
 Db 1737 CAACGATGTGTGTCTTCTTTATTTACACAGACTGATGACCGAAGAAAGATCGATGTCA 1796  
 RESULT 5  
 AX090345 1942 bp DNA linear PAT 21-MAR-2001  
 LOCUS AX090345  
 DEFINITION Sequence 38 from Patent WO0116308.  
 ACCESSION AX090345  
 VERSION AX090345.1 GI:13444207  
 KEYWORDS  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 REFERENCE  
 1 Lasserre M. and van Eenennaam A.  
 AUTHORS Plant sterol acyltransferases  
 TITLE Patent: WO 0116308-A 38 08-MAR-2001;  
 JOURNAL MONSANTO COMPANY (US)  
 FEATURES  
 source location/Qualifiers  
 1..1942  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3702"  
 BASE COUNT 461 a 421 c 425 g 635 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 5..06e-222 Length: 1942  
 Score: 2771.00 Matches: 520  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0  
 US-09-623-514a-2 (1-520) x AX090345 (1-1942)  
 QY 1 MetAlaIleLeuAspSerAlaGlyValIleThrValIleThrGluAsnGlyGlyGluPhe 20  
 |||||  
 Db 237 ATGGCATTTTGGATTTCTGCTGCGCTTACTACGGTACGAGAACGTTGGCGAGAGTTTC 296  
 QY 21 ValAspLeuAspArgLeuArgArgArgIysSerArgSerAspSerSerAsnGlyLeuLeu 40  
 |||||  
 Db 297 GTTCGATCTTGATAGCGTTCTGTCAGCGAAATCGAATTCGATTTCTTAACGACTTCTT 356  
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60  
 |||||  
 Db 357 CTCTCGTTCGCGATTAATTAATTCCTTCGAGATGTTGGAGCTCCCGCAGCTTAGG 416  
 QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
 |||||  
 Db 417 GATCGGATTTGATTCCTTTTAACGATGACGCTCAGGAAAGCCAAATTTGCCCGAGAT 476  
 QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyArgGlyGlyGlyGlyArgGlyAsn 100  
 |||||  
 Db 477 AATTAACGGTGGTGGCATATAAAGCGTGTGGAAGAGGGCGGCAAGAAAGAGAAAC 536  
 QY 101 AlaaspaIatPhPheThyTyrArpProSerValProAlaHisArgAlaArgIuSer 120  
 |||||  
 Db 537 GCCGATGTCACTTTACATGACCGCTGGTTCAGCTCATCGAGGCGAGAGAGACT 596  
 QY 121 ProLeuSerSerAspAlaIlePheIysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
 |||||  
 Db 597 CCACCTAGCTCCGACCAATCTTCAACAGACGCAATGCCGATTTATTCAACTCTGTGA 656

```

Db      |||
259 CTCTGGTCCGAAATAATTCCTTCGGATGATGTTGGAGCTCCGCGACCTTGG 318
Oy      |||
61 AspargileaspservalevalasnaspsalaglnclythrAlaasoleualaglasp 80
Db      |||
319 GATCGGATGTAATCCGCTTGTAACGATGACCTCGAGGACAGCAATTTGGCCGGAAT 378
Oy      |||
81 Asnansglylglyaspasnasnglylglyargglylglygluglyargglyasn 100
Db      |||
379 AATAACGGTGGTGGCGATAATACGGTGGGAAGAGCGCGGAGGAAGGAAGGAAC 438
Oy      |||
101 AlaspsalathrphethrThrArgProserValProAlaHisArgArgAlaArgLuser 120
Db      |||
439 GCCGATGCTACGTTTACGATATCGACCGCTGCTCCAGCTCATCGAGGCGGAGAGAGT 498
Oy      |||
121 ProleuserSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeucysVal 140
Db      |||
499 CCACCTACCTCGACGCAATCTTCAACAGAGCCATGCCGATTTATTAACCTCTGTGTA 558
Oy      |||
141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrp 160
Db      |||
559 GTAGTCTTATGTCGTAAACAGTACTCATCATCGAAATCTTATGAGATGTTGG 618
Oy      |||
161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProleuPheMet 180
Db      |||
619 TTGATCAGACGAGATTCCTGCTTGAAGTCAAGATCGCTGCGAGATGGCCCTTTTCATG 678
Oy      |||
181 CysCysIleSerLeuSerIlePheProleuAlaAlaPheThrValGlyLysLeuValLeu 200
Db      |||
679 TGTGTATATCCCTTCGATCTTTCCTTGGCTGCTTACGCTTGAAGAAATGGTACTT 738
Oy      |||
201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
Db      |||
739 CAGAAATACATATCAGAACCTGTGTGTCATCTTCTTATATATATATACCATGACAG 798
Oy      |||
221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
Db      |||
799 GTTTGTATCCAGTTTACGTACCCCTAAGGTGATGCTGCTTTTTCAGGTCTACT 858
Oy      |||
241 LeuMetLeuLeuThrCysIleValTyrPheLysLeuValSerTyrAlaHisThrSerTyr 260
Db      |||
859 TTGATGCTCCCTACCTGATGCTGTGCTAAAGTTGCTTCTTATGCTCATACATGCTAT 918
Oy      |||
261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyrVal 280
Db      |||
919 GACATAGATCCCTAGCCCATGACAGCTGATGAAGCCATCTCTGAAGTCTCTACTACGTT 978
Oy      |||
281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
Db      |||
979 AGCTTGAAGACCTTGGCATATTTCAATGCTGCCCTCCACATGCTTATCACGCCAAGTAT 1038
Oy      |||
301 ProArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIle 320
Db      |||
1039 CCAGCTTTCGATGATGATACGGAAGGTTGGCTGCTGCTCAATTTGCCAAACTGCTCTA 1098
Oy      |||
321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340
Db      |||
1099 TTCACCGGATTCATGCGGATTTATATAGAACATATATATATCCATATGTCAGAGAAC 1158
Oy      |||
341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360
Db      |||
1159 AAGCATCTTGAAGGCGATCTTATATGCTATGTAAGAGAGCTTCAACCTTTCACCTT 1218
Oy      |||
361 ProAsnLeuTyrValTyrPheLysCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380
Db      |||
1219 CCAATTTATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278
Oy      |||
381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLys 400
Db      |||
1279 TTGGCAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
Oy      |||
401 SerValGlyAspTyrTyrPheMetTrpAsnMetProValHisLysTyrMetValArgHis 420

```

```

Db      |||
1339 AGTGTGGAGATTACTGGAGAAATGTGGAATATGCCCTGTTTCATTAATGAGTGTTCGACAT 1398
Oy      |||
421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
Db      |||
1399 ATATATCTTCCGCTGCTTGGCAGCAAGATACCAAGACATCTGCCATTTATCATTTGCTTTC 1458
Oy      |||
441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
Db      |||
1459 CTAGCTCTGCAGATCTTTCATGAGCTATGATGATGATGATGATGATGATGATGATGATGAT 1518
Oy      |||
461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValIlePheIleThrAsnTyrLeuGln 480
Db      |||
1519 TGGGCTTTTCTTGGGATTAATGTTTCAGGTGCTTGGCTTGTCTTCATCACAACTATCTAC 1578
Oy      |||
481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
Db      |||
1579 GAAAGTTTGGCTCAACGCTGGGGAACATGATCTTCTGTGTCATCTTGTGATTTTCGGA 1638
Oy      |||
501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
Db      |||
1639 CAACGATGTGTGCTCTTATATACACGACCTGATGAACCGAAAGATGATGTCA 1698

```

## RESULT 4

```

LOCUS      AR227739      1942 bp      DNA      linear      PAT 20-DEC-2002
DEFINITION Sequence 1 from patent US 6444876.
ACCESSION  AR227739
VERSION     AR227739.1 GI:27266330
KEYWORDS
SOURCE
ORGANISM   Unknown.
            Unclassified.
REFERENCE  1 (bases 1 to 1942)
AUTHORS   Lassner, M. W. and Ruzelinsky, D. M.
TITLE     Acyl CoA: cholesterol acyltransferase related nucleic acid
SEQUENCES
JOURNAL    Patent: US 6444876-A 1 03-SEP-2002;
FEATURES   Location/Qualifiers
            source          1..1942
            BASE COUNT      461 a 421 c 425 g 635 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      5,06e-222      Length:      1942
Score:          2771.00      Matches:      520
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:     100.00%      Indels:      0
DB:              Gaps:      0

```

```

US-09-623-514A-2 (1-520) x AR227739 (1-1942)
Oy      |||
1 MetalAlaIleLeuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
Db      |||
237 ATGGCGATTTTGGATTCTGCTGCGCTTACTACGCTGACGAGGAACGGTGGCGGAGAGTTC 296
Oy      |||
21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40
Db      |||
297 GTCGATCTTGATPAGGCTTCTGTCAGAGGAATCGAGATGCAATCTTCTTCAACGCACTTCTT 356
Oy      |||
41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
Db      |||
357 CTCTCTGCTTCGCAATAATAATTCCTTCCGATATATGTTGAAGCTCCCGCGAGACTTGG 416
Oy      |||
61 AspargileaspservalevalasnaspsalaglnclythrAlaasoleualaglasp 80
Db      |||
417 GATCGGATGTAATCCGCTTGTAACGATGACCTCGAGGAACGCCAATTTGGCCGAGAT 476
Oy      |||
81 Asnansglylglyaspasnasnglylglyargglylglygluglyargglyasn 100
Db      |||
477 AATAACGGTGGTGGCGATTAATACGGTGGGAAGAGCGCGGAGGAAGGAAGGAAGAAC 536

```

[illegible]

Oy	501	GINPROMETCYSVALLEULEUTYTYRTHISASPLEUNEATASNAATGLYSGLYSERMETSER	520
Db	1588	CAACCGATGTGTGCTCTTTATTATACACACACTGATGACCAACCAAAAAGATCATGTCA	1647
RESULT 3			
LOCUS	ATH238008	1904 bp	linear
DEFINITION	Arabidopsis thaliana mRNA for diacylglycerol acyltransferase.		
ACCESSION	AJ238008		
VERSION	AJ238008.1	GI:5123717	
KEYWORDS	Dgat gene; diacylglycerol acyltransferase.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
AUTHORS	1		
TITLE	Zou, J., Wei, Y., Jako, C., Selvaraj, G. and Taylor, D.C.		
JOURNAL	The Arabidopsis thaliana TAG1 gene encodes for a diacylglycerol acyltransferase		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 1904)		
TITLE	Zou, J.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (06-APR-1999) Zou J., Plant Biotechnology Institute, National Research Council of Canada, 57H 0M9, Saskatchewan, CANADA		
FEATURES	Location/Qualifiers		
source	1..1904		
gene	/organism="Arabidopsis thaliana"		
	/mol_type="mRNA"		
	/db_xref="taxon:3702"		
	1..1904		
	/gene="Dgat"		
	139..1701		
	/gene="Dgat"		
	/function="lipid synthesis"		
	/codon_start=1		
	/product="diacylglycerol acyltransferase"		
	/protein_id="CA845373.1"		
	/protein_xref="GI:5123718"		
	/db_xref="GI:5123718"		
	/translation="MALIDSAGVTTTENGEGEVDIDLRLRRKSRSDSSNGLLSGSSDNNSPDSVGVGADVRDIRIDSVNDADOGNANLADNGNNGDNGGRRGGEGEGCNDATFPYRSPVPAHRARARESPSSDAIFQSHAGLELVCVVYLIAVNSRLILENLMKYGMLIRDFDEWSSRLSDWPLFMCCISLSIEPLAATVEKLYLOKYSSEVVEVLEIIIMTEVLYPVVTLRCDSAFLSGVTLMLTLNLCIYWLKATSVYAHNSYDIRSLANADANPEVSVYSIKSLATFMYAPFLCYCPSPYRSCIRKGVNAQPAKLYITFGPKGIIIEQIYNPIVNRKHPLEKGLDLYIAERLYKLSVPLVLYWLCMEYCFHLMNIIIAELLCFEDREFEKDWNAKSVGSDYWRMMNMVPAHKMVRHRIYFPCLRSKIPKTLAIIIFLVSAYFHEICIAVCRLEKTLMAFLGIMFQVPLVEITVTLDERFGFSITVGNMIFWIFICIFGQPMCVLLYLYHDLNLRKSGSKS"		
BASE COUNT	458 a	399 c	431 g
ORIGIN			616 t
Alignment Scores:			
Pred. No.:	4.92e-222	Length:	1904
Score:	2771.00	Matches:	520
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0
US-09-623-514A-2 (1-520) x ATH238008 (1-1904)			
Oy	1	MethaHalleuapSerAlaglyValThrValThrgluanglyglygluPhe	20
Db	139	ATGGCGATTTGGATCTGCTGGGGCTTACCTACGGGTGACGGAGACGGTGGCGAGAGTTC	198
Oy	21	ValaspleuAaPAqyleuAaRGAARqlySerAqySerAaPpaSerSerAaNglyLeuLeu	40
Db	199	GTCATCTTTAAGCGCTTCGTCGACGGAAATCGAGATCGGATTTCTTACGCGACTTCTT	258
Oy	41	LeuserglySerAaPaMaPaMaPaMaPaMaPaMaPaMaPaMaPaMaPaMaPaMaPaMaPaMa	60

Alignment Scores:			
Pred. No.:	4,92e-222	Length:	1904
Score:	2771.00	Matches:	520
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	8	Gaps:	0
US-09-623-514A-2 (1-520) x ATH238008 (1-1904)			
QY	1 Metalalieleuaspseralaglyalthrvalthrgluansglyglygluphe	20	
DB	139 ATGGCGATTTGGATTCTGCTGGGGTTCCTCGGTGACGGGAACGGTGGCGGAGACTTC	198	
QY	21 Valaspleuaspargyleuargarglyssertrgserspsserseransglyleu	40	
DB	199 GTCGATCTTGTAAGGCTTCGTGCGACGGAAATCGAGATCGGATTTCTTACGGACTTCTT	258	
QY	41 leuSerglySersaspasnasnserserProSersaspvalglalaproalaaspvalarg	60	

Db	841	AGCTTGAAGAGCTTGGCATATTTCATGGTCCGTCACATTTGTGTTATCAGCAAGTTAT	900
QY	301	ProArigSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle	320
Db	901	CCACGTTCTGCATGTATATACGGAAGGGTTGGGGGCGCTGCATTAATTTGCAAAACTGGTCATA	960
QY	321	PheThrArgLysPheMetGlyPheIleIleGlnGlnTrpIleAsnProIleValArgAsnSer	340
Db	961	TTACACCGGATTCATGGGATTTTATATATTAACCAATATATTAATTCCTATTGTGATCGGAACCA	1020
QY	341	LysHisProLeuLysGlyAspLeuLeuTrpValAlaIleGluArgValLeuLysLeuSerVal	360
Db	1021	AAGCATCTCTTTGAAAGGCGCATCTTCTATATGCTATTTGAAGAGTGTAAAGCTTTCAAGTT	1080
QY	361	ProAsnLeuTrpValTrpLeuLysMetPheTrpCysPhePheHisLeuTrpLeuAsnIle	380
Db	1081	CCAAATTTATATGTGTGGCTCTGCATGTTCACCTGCTCTTCCACCTTTGGTTAAACATA	1140
QY	381	LeuAlaGlnLeuLeuLysCysPheGlyAspArgGlnPheTrpLysAspTrpTrpAsnAlaLys	400
Db	1141	TTGGCAGAGCTTCTCTGCTGGGGCATGTGAATTCCTCAAAAGATGGTGATGCCAAA	1200
QY	401	SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis	420
Db	1201	AGTGTGGGAGATTACTGCAGATGTGGAAATATGCTGTTCATTAATTAATGATGGTTCGACAT	1260
QY	421	IleTrpPheProCysLeuArgSerLysIleProLysTrpLeuAlaIleIleIleAlaPhe	440
Db	1261	ATATACCTTCCCGTGGCTTCGCGACGACCAATACCAAAAGACATCGCCATATATCTGCTTTC	1320
QY	441	LeuValSerAlaValAlaPheHisGlnLeuLysIleAlaValProCysArgLeuPheLysLeu	460
Db	1321	CTAGCTCTGCAAGCTTTTCATGAGGTAAAGCATCGCATGCTCTGCTCTTCAAGCTA	1380
QY	461	TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTrpLeuGln	480
Db	1381	TGGCTTTTCTTTGGGATTATATGTTTCAGGTGCTTTGGTCTTCATCACAAACTATCAAG	1440
QY	481	GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly	500
Db	1441	GAAAGCTTTGGCTCAACGCTGGGGAACATGATCTTCGTTCATCTTCGATTTTCGGA	1500
QY	501	GlnProMetCysValLeuLeuTrpTrpHisAspLeuMetAsnArgLysGlySerMetSer	520
Db	1501	CACCGCATGTGTGCTCTTCTTATATACACGACCTGATGAACCGAAAGATCGATGCA	1560
RESULT 2			
AF051849		1845 bp	mRNA
LOCUS			linear
DEFINITION	Arabidopsis thaliana diacylglycerol acyltransferase (DAGAT) mRNA,		PLN 24-JAN-2000
ACCESSION	AF051849		complete cds.
VERSION	AF051849.1		GI:6625552
KEYWORDS			
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.		
	1 (bases 1 to 1845)		
REFERENCE	Bouvier-Herve, P., Benveniste, P., Oelkers, P., Sturley, S.L. and		
AUTHORS	Schaller, H.		
TITLE	Expression in yeast and tobacco of plant cDNAs encoding acyl		
JOURNAL	CoA:diacylglycerol acyltransferase		
MEDLINE	Eur. J. Biochem. 267.(1), 85-96 (2000)		
PUBMED	20069349		
REFERENCE	10601854		
AUTHORS	Benveniste, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (28-Feb-1998) Institut de Biologie Moleculaire des		
	Plantes, Centre National de la Recherche Scientifique, 28 rue		

COMMENT	Goethe, Strasbourg 67083, France
On Jan 24, 2000 this sequence version replaced gt:6650202.	
FEATURES	Location/Qualifiers
source	1. .1845
gene	/organism="Arabidopsis thaliana"
CDS	/mol_type="mRNA"
	/cultivar="Columbia"
	/db_xref="taxon:3702"
	/tissue_type="hypocotyl"
	/clone_1ib="ABRC library, catalog number CD4-15"
	/dev_stage="3 day-old seedling"
	/note="screened with EST, Genbank accession Number AA042298"
	1. .1845
	/gene="DAGAR"
	88. .1650
	/gene="DAGAR"
	/function="catalyzes the acylation of diacylglycerol by oleoyl-coA to give triacylglycerol"
	/note="acyl-CoA cholesterol acyltransferase homolog"
	/codon_start=1
	/product="diacylglycerol acyltransferase"
	/protein_id="AAF19262.1"
	/db_xref="GI:6625553"
	/translation="MAIDSDAGVTYVTEGGEFVDLDLRRRKSSDSSNGLLSGD DNSSDDVGADPADVRDRIYDVNDQSTAMLAGNNNGDNNGGRGGERGNNA ATFYRSPVPAHRRARESPLESDDAIFGQAGFNCVAVLIVANNRIENIMKYG LIRDFEFSRSRLRMPLEMCISLIFLAFTYERLVLOKYSPPVYFLHITITTE TELVIPVYTLERCDSEFLSGVTMLITFCIVMKLYVATSDVIRSLANAADKPE SYVSLKSLAYFMVAPTLCYOPSPYSCIRGQVARNOPAKIVIFPGFITEQYI PIVRNSKHPLEKGDILXAIRVYIKLSIPNUTYWCMTYCFPHMLNLALICGDBE YKDMNNAKSVGDYWRMNMNPHKMWVRHRIYFQLRKRITPTLAIILAFVSAFHE IAVPRFLKMAFLGIMFQVPLVFITNYLQERFRGTVGNMFIKIFGIPGCVLL YHDLNLRKGSMS"
BASE COUNT	453 a 378 c 421 g 593 t
ORIGIN	
Alignment Scores:	
Pred. No.:	4.72e-222 Length: 1845
Score:	2771.00 Matches: 520
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
	Gaps: 0
US-09-623-514A-2 (1-520) x AF051849 (1-1845)	
QY	1 Metalaieleuaspserlaaglyvalmrtrhvalthgluansnelygllyglupne 20
DB	88 ATGGCGATTTCGATTTCCTCGCTACTACGGTACGAGAAACGTCGGCGAGAGTTTC 147
QY	21 VALAASPLAASPAARGLEUARGTARGTGLYSERARGSERASPOSERASNGLYLEUEN 40
DB	148 GTCCGATCTTGTAAGGCTTCGTGCGACGGAATGCAATCGGATCTTCTTAACGACTTCTT 207
QY	41 LeuSergIySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
DB	208 CTCTCTGGTTCGGAATATATCTCTCTCGATGATGTGGAGCTCCGCGCAGCTTAGG 267
QY	61 AsparGlyLeuAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
DB	268 GATTCGATTGATTCGCTGTTAACGATGACGCTCAGGAAACAGCCAAATTTGGCCGGAGAT 327
QY	81 AsnaAngLyGLyGLyAspAsnAsnGlyGLyGLyArgGLyGLyGLyGLyGLyAsn 100
DB	328 AATAACGGTGGCGGCGATATATACGGTGTGCAAGACGGCGCGAGACAGAGAAAC 387
QY	101 AlaAspAlaThrPheThrTyraArgProSerValProAlaHisArgArgAlaArgGluSer 120
DB	388 GCCGATGCTACGTTTACGATGACGACGCTGCTCCAGCTCATTGCGAGGGCGAGAGAGAG 447
QY	121 ProLeuSerSerAspAlaAlaPheIleYsGlnSerHisAlaGlyLeuPheAsnLeuCyVal 140

BT008883  
LOCUS 1563 bp mRNA linear PLN 13-JUN-2003  
DEFINITION Arabidopsis thaliana At2g19450 mRNA, complete cds.  
ACCESSION BT008883  
VERSION 1  
KEYWORDS GI:31711931  
SOURCE FLI CDNA  
ORGANISM Arabidopsis thaliana (thale cress)  
Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 1563)  
REFERENCE  
AUTHORS Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Bowser, L., Carninci, P.,  
Dale, J.M., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,  
Karlin-Neumann, G., Kawai, J., Lam, B., Lin, J., Miranda, M.,  
Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L.,  
Sekurai, T., Satou, M., Seki, M., Southwick, A., Toriumi, M., Wong, C.,  
Wu, H.C., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W., Theologis, A.,  
and Ecker, J.R.  
Arabidopsis ORF clones  
Unpublished  
2 (bases 1 to 1563)  
Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Bowser, L., Carninci, P.,  
Dale, J.M., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A.,  
Karlin-Neumann, G., Kawai, J., Lam, B., Lin, J., Miranda, M.,  
Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L.,  
Sekurai, T., Satou, M., Seki, M., Southwick, A., Toriumi, M., Wong, C.,  
Wu, H.C., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W., Theologis, A.,  
and Ecker, J.R.  
Direct Submission  
Submitted (13-JUN-2003) Salk Institute Genomic Analysis Laboratory  
(SIGNAL), Plant Biology Laboratory, The Salk Institute for  
Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037,  
USA  
RIKEN Genomic Sciences Center (GSC) members carried out the  
collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN  
Arabidopsis Full-Length cDNA'): Seki, M., Narusaka, M., Ishida, J.,  
Satou, M., Kamiya, A., Sekurai, T., Carninci, P., Kawai, J.,  
Hayashizaki, Y., and Shinozaki, K.  
The Salk, Stanford, PGSC (SSP) Consortium members constructed and  
sequenced the PUNI (ORF) clones using the RAPL cDNAs: Kim, C.J.,  
Chen, H., Cheuk, R., Shinn, P., Bowser, L., Chan, M.M., Chang, C.M.,  
Dale, J.M., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J.,  
Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L.,  
Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K.,  
Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.  
Kim, C.J. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to  
this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk)  
contributed equally to this work as PIs.  
Location/Qualifiers  
1. 1563  
/organism="Arabidopsis thaliana"  
/mol\_type="mRNA"  
/db\_xref="taxon:3702"  
/chromosome="2"  
/clone="U16373"  
/note="This clone is in PUNI 51  
ecotype: Columbia"  
1. 1563  
/note="dlaacylglycerol O-acyltransferase"  
/codon\_start=1  
/product="At2g19450"  
/protein\_id="AAP68322.1"  
/db\_xref="GI:31711932"  
/translation="MALIDSAGVTYTTENGGEFVLDRLRRKSRSDSSNGLLSGS  
DNNSPDVGAPADVDKRDIDSYVNDHOGANLANLGGDNGGNGGEGEGEGND  
ATFTYRSPVAHRRARSPSSDAIFKQSLAAGFNLCAVYLIVNSRLILEMKGW  
LITPDYRSPVARRRSPSSDAIFKQSLAAGFNLCAVYLIVNSRLILEMKGW  
TEVLYPVYTLRCDSAFLSGVTMLTLCIKVSAHRTSYAHTSYLANADANBYE  
SYVSLKSLAYFWAPVAPICOPSPRSACIRKGVAROFKATYIFGFGGFTIEQYIN  
PIYRNSKHPLKGLLVAIERVLAISPVNLTVMLCATYCFPHMLNTLAIELCIGDRE

YKDMNNAKSVGDYWRMMNMPVHKMMVRRHIFPCILRSKIPRTALIIIAFLVSAVHELC  
IAVCRLEFLMAFLGIMFQVPLVFTNYLQERGSYVNMIFEFICIFQPMCVLLY  
YHDLNMRKSGMS"  
BASE COUNT 376 a 322 c 371 g 494 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,78e-222 Length: 1563  
Score: 2771.00 Matches: 520  
Percent Similarity: 100.00% Conservative: 0  
Best local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 8 Gaps: 0  
US-09-623-514a-2 (1-520) x BT008883 (1-1563)  
OY 1 MetAlaIleleuSpSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20  
|||||  
DB 1 ATGGCGATTTTGGATTTCTGCTGGCGTTACGCGTGACGAGAACGGTGGCGAGATTC 60  
21 ValAspLeuAspArgLeuArgArgGlySerArgSerAspSerAspSerAspLeu 40  
61 GTGATCTTATAGCGCTTCGTCGACGGAATCGAGATCGATTCCTTAACGACTTCT 120  
OY 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60  
121 CTTCTCTGTTCCCATATATATTTCTCCCTCGAGATGATGTGAGACTCCCGCGAGTTAG 180  
OY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
|||||  
DB 181 GATCGATTAATTCCTGTTACGATGAGCGTCGAGGAACGCAATTTGGCGGAGAT 240  
81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGly 100  
241 AATTAACGCTGCTGCGCATATTAACGCTGCTGCGAAGACCGCGGAGAGAGAGAAC 300  
OY 101 AlaAspAlaThrPheThrTyArgProSerValProAlaHisArgAlaArgGluSer 120  
301 GCCGATCTTACGATTAATTCGATTCACCGCTTCGATTCATCGAGCGGAGAGAGAT 360  
121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140  
|||||  
DB 361 CCACCTTACCTCCGACGGAATCTTCMAACAGACCATCCGATTAATCAACCTCTGTGA 420  
141 ValValLeuLeuAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyArgLys 160  
421 GTAGTTCTTATGCTGTAACAGACGATCATCATCAAAATCTTATGAGATGTGCTG 480  
OY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180  
481 TTGATCAGAACGAGATTTCTGTTAGTTCAAGATCCGTCGAGATTGGCCGCTTTTCATG 540  
OY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGlyLysLeuValIleu 200  
541 TGTGTTATATCCCTTTTCGATCTTCTTCCTTGCGTCCCTTACGGTTGAGAAATGATCTT 600  
OY 201 GlnLysTyrlIleSerGluProValIleIlePheLeuHisIleIleIleThrMetThrGlu 220  
601 CAGAAATACATATCAGAACCTGTTGATCTTCTTCTTCTTATTTATTCACATCAGACAG 660  
OY 221 ValLeuTyProValTyValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
661 GTTTTGTATCCAGCTTACGTCACCCATGAGTGTGATCTCTTTTATATAGGTGTACAC 720  
DB 241 LeuMetLeuLeuThrCysIleValITPLeuLysLeuValSerTyrlAlaHisTrpSerTyrl 260  
721 TTATGCTCTCACTTGCATTTGTGTGCTAAAGTGTGTTCTTATGCTCACTTACTACTAT 780  
OY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProLysValSerTyrlTyrlVal 280  
781 GACATTAAGATCCCTAGCAGATGAGCGATGAAGCCCATCTGAAGTCTCTACTACGTT 840  
OY 281 SerLeuLysSerLeuAlaTyrlPheMetValAlaProThrLeuGlySerGlnProSerTyrl 300

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 31, 2003, 05:01:50 ; Search time 4675 Seconds  
(without alignments)  
4550.376 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771  
Sequence: 1 MAILDSAGVTVTENGGEF.....OPMCVLYYHDMKRKGSMS 520

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+pn.model -DEV=xlP  
-Q/cgn2.1/USPTO.spool/US09623514/runat.29082003.152159.54/app.query.fasta.1.711  
-DB=GenEmbl -QFMT=fastp -SUFFIX=rge -MINMATCH=0.1 -IOFCL=0 -IOFEXT=0  
-UNIT=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09623514.ecgn.1.1.5265.etrnat.29082003.152159.54 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_da:\*  
2: gb\_hg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_da:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vl:\*  
30: em\_hg\_hum:\*  
31: em\_hg\_inv:\*  
32: em\_hg\_other:\*  
33: em\_hg\_mus:\*  
34: em\_hg\_pin:\*  
35: em\_hg\_rtd:\*  
36: em\_hg\_mam:\*  
37: em\_hg\_vrt:\*  
38: em\_sy:\*  
39: em\_hgo\_hum:\*  
40: em\_hgo\_mus:\*  
41: em\_hgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	1563	8 BT008883	BT008883 Arabidops
2	2771	100.0	1845	8 AF051849	AF051849 Arabidops
3	2771	100.0	1904	8 ATR238008	ATR238008 Arabidops
4	2771	100.0	1942	6 AR227739	AR227739 Sequence
5	2771	100.0	1942	6 AX090345	AX090345 Sequence
6	2771	100.0	1942	6 AX090349	AX090349 Sequence
7	2771	100.0	1988	6 ATR131831	ATR131831 Arabidops
8	2771	100.0	2005	8 AY054480	AY054480 Arabidops
9	2351.5	84.9	1537	8 AF251794	AF251794 Brassica
10	2309.5	83.3	1512	8 AF164434	AF164434 Arabidops
11	2225	80.3	92822	8 AC005917	AC005917 Arabidops
12	2225	80.3	97495	8 AC003058	AC003058 Arabidops
13	1880	67.8	2090	8 AT084052	AT084052 Tropaeolu
14	1854	66.9	1964	8 AF298815	AF298815 Perilla f
15	1842.5	66.5	2099	8 AF129003	AF129003 Nicotiana
16	1833	66.1	1446	8 AF155224	AF155224 Brassica
17	813	29.3	43490	8 AP006408	AP006408 Lotus jdp
18	794	28.7	1976	9 AF059202	AF059202 Homo sapi
19	784	28.7	1998	9 BC015762	BC015762 Homo sapi
20	793	28.6	2000	9 BC023565	BC023565 Homo sapi
21	791.5	28.6	1518	9 AB057815	AB057815 Homo sapi
22	789	28.5	1856	3 AK114382	AK114382 Clona int
23	773.5	27.9	1661	10 AB062763	AB062763 Rattus no
24	773.5	27.9	1681	10 AB062759	AB062759 Rattus no
25	773.5	27.9	1681	10 AB062760	AB062760 Rattus no
26	773.5	27.9	1681	10 AB062761	AB062761 Rattus no
27	773.5	27.9	1681	10 AB062762	AB062762 Rattus no
28	773.5	27.9	1766	6 AR227752	AR227752 Sequence
29	773.5	27.9	1766	6 AR227753	AR227753 Sequence
30	773.5	27.9	1766	6 AX090340	AX090340 Sequence
31	772.5	27.9	1792	9 AF236018	AF236018 Sequence
32	771.5	27.8	1547	10 AB057816	AB057816 Mus muscu
33	771.5	27.8	1650	10 AF078752	AF078752 Mus muscu
34	771.5	27.8	1776	10 BC003717	BC003717 Mus muscu
35	762.5	27.5	1751	10 AP296131	AP296131 Rattus no
36	751.5	27.1	1935	4 AY093657	AY093657 Sus scrof
37	735.5	26.5	1493	9 BC006263	BC006263 Homo sapi
38	725.5	26.2	2074	3 AY051835	AY051835 Drosophill
39	724.5	26.1	1497	3 AF221132	AF221132 Caenorhab
40	720.5	26.0	2117	3 AF468650	AF468650 Drosophill
41	720.5	26.0	2342	3 AF468649	AF468649 Drosophill
42	709.5	25.6	1895	6 AR227751	AR227751 Sequence
43	709.5	25.6	1895	6 AX090339	AX090339 Sequence
44	696	25.1	185932	2 AP003714	AP003714 Oryza sat
45	622.5	22.5	34948	3 CBG333E23	CBG333E23 Caenorhab

RESULT 1

## ALIGNMENTS



640	CGCCCCAATGCGTTATATGCGGCCAAATGCTTTTGTGTGGCTGCATTCACGAGTTGAGAA	699
Qy	197 sLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleth	217
Db	700 GCGCCTGGCGTGGGTGCGCTGACGGAGAGCGGGACTGCTGTCACAGTACCCAACT	759
Qy	217 rMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSe	237
Db	760 GGCCACCATTCTGTGTTCCTCCAGCGGTGTGGTCTTACTGGTTAGTCTATCACTCCAGT	819
Qy	237 rGlyValThrLeuMetLeuLeuThr--CysIleValThrLeuLysLeuValSerTyrAl	256
Db	820 GGGTCCTCCTGCTGGCGTGAAGGGCACACCATCTCTTCCTCAAGCTCTTCCTACCG	879
Qy	256 aHisThrSerTyrAspIleArgSer-----LeuAlaAsnAl	268
Db	880 C-----GAGCTCAACTCATGTGCGCGAGGGCCAGGCCAAGGCTGCCCTGTC	927
Qy	268 aAlaAspLysAlaAsnPro-----GluValSerTyr-----TyrValSe	281
Db	928 AGGGAAGAGAGCCAGCAGTAGTCTGTCGCCCGCACACCGTGAGCTACCGGACAACTGAC	987
Qy	281 rLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrPr	301
Db	988 CTACCGCGGATCTCTACTACTCTCTTCGCGCCGCCACCTGTGTGTACGAGTCACTTCC	1047
Qy	301 oArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIlePh	321
Db	1048 CCGCTCTCCCGCATCGGAAGCGCTTTCTGCTGCGAGCGGATCCTTGAGATGCTGTCTT	1107
Qy	321 eThrGlyPheMetGlyPheIleIleGlnGlnTyrIleAsnProIleValArgAsnSerLy	341
Db	1108 CACCCAGCTCCAGTGGGCGTGAATCCAGCAGTGGATGGTCCCCACCATCCAGAACTCCAT	1167
Qy	341 sHisProLeuLysGlyAspLeuLeuTyrAla-----IleGluArgValLeuLysLeuSe	359
Db	1168 GAAGCCCTTCAAG---GACATGGACTACTCACGCATCATCGAGCGCTCCTGAAGCTGGC	1224
Qy	359 rValProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAs	379
Db	1225 GGTCGCCCAATCACTCATCTGGGTGCATCTTCTTCTACTGGCTCTCCATCTCCGCTGAA	1284
Qy	379 nIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAl	399
Db	1285 TGCGTGGCTGAGCTCATGTCAGTTTGAGACCGGGAGTCTACCGGACTGGTGGAACTC	1344
Qy	399 aLysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValAr	419
Db	1345 CGAGTCTGTCACTTCTTGGCAGAACTGGAAATCCCTGTGCACAAGTGGTGCATCAG	1404
Qy	419 gHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAl	439
Db	1405 ACATCTCTACAGCCCATGCTTCGAGGGGCACGACGAGTGGATGCCAGGACAGGGGT	1464
Qy	439 aPheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLy	459
Db	1465 GTTCTTGGCCTCGGCTTCTTCCACGAGTACCTGGTGGCGTCCCTCTCGAATGTTCCG	1524
Qy	459 sLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTy	478
Db	1525 CCTCTGGGCTTTTCCAGGGATGATGCTCAGATCCCACTGGCTGTGTCTGGCGGCGCTT	1584
Qy	478 rLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPhePheCysTrl	498
Db	1585 TTTCCAGGCCAACTAT-----GGCAACGCGAGCTGTGTGG---CTGTGCGTCAAT	1629
Qy	498 ePheGlyGlnProMetCysValLeuLeuTyrTyrHisAsp	511
Db	1630 CATCGACGCCAATAGCGCTCTCATGTAGTGCACGAC	1669



QY 366 TrpLeuCysMetPheTyrCysPheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeu 385  
 Db 391 TGGCTTTCATGCTTATCTTTTCATTTTGGCTGAACATTTAGCTTAACTCCTC 450  
 QY 386 CysPheGlyAspArgGluPheTyrLysAspTyrPheAsnAlaLysSerValGlyAspTyr 405  
 Db 451 TGTTTCGTCACCGTGAACTTCTATAAGGAGCTGGTGAATGCAAACTGTTGAAGAGTAC 510  
 QY 406 TrpArgMetTrpAsnMetProValHisLysLysTrpMetValArgHisLysLysPheProCys 425  
 Db 511 TGGAGGATGTGGAAACATGCTGTTCATPAGTGGATCATCAGACACATATATTTCATGT 570  
 QY 426 LeuArgSerLysIleProLysThrLeuAlaIleIleLeuAlaPheLeuValSerAlaVal 445  
 Db 571 ATAGGNNAGGCTTCCAGGGGTAGCTATTCTTAATCTCGTTTGGTTTTCAGCTGTA 630  
 QY 446 PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly 465  
 Db 631 TTCCATGAGATATGATTGCGGTGCGGCCACATTTTCAANTCTCGGCATTTCTCGGG 690  
 QY 466 IleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySer 485  
 Db 691 ATCATGTTTCAGATACCGTTGGTATTCTTGACAAAGATATCTCCATGCTAGTTCAAGCAT 750  
 QY 486 Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCys 504  
 Db 751 GTAATGTGGGCAACATGATATTGTTGCTC---TTCAGTATAGTCCGACAGCCGATGNG 807  
 QY 505 ValLeuLeuTyrTyrHisAspLeuMet-AsnArgLysGlySerMetSer 520  
 Db 808 TGTCTCTATAACTAAACATGACGTATGAAACAAAGCCCAAGCAAGT 856  
 RESULT 15  
 AAA76169  
 ID AAA76169 standard; DNA; 1976 BP.  
 AC AAA76169;  
 XX  
 DT 14-DEC-2000 (first entry)  
 XX  
 DE Human ACAT Related Gene Product 1 ARGPI coding sequence.  
 XX  
 KW Human; ACAT Related Gene Product 1; ARGPI; gene therapy; enzyme;  
 KW acyl Coenzyme A-cholesterol acyltransferase 1; ACAT1;  
 KW sterol esterification; lipid homeostasis; diacylglycerol acyltransferase;  
 KW DGAT; cholesterol; triglyceride biosynthesis; hypertriglyceridaemia;  
 KW hyperlipidaemia; atherosclerosis; heart disease; obesity; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 245..1711  
 FT /\*tag= a  
 FT /product= "Human ARGPI"  
 XX  
 PN US6100077-A.  
 XX  
 PD 08-AUG-2000.  
 XX  
 PF 01-OCT-1998; 98US-0165042.  
 XX  
 PR 01-OCT-1998; 98US-0165042.  
 XX  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 XX  
 PI Sturley SL, Oelkers P;  
 XX  
 DR WPI; 2000-557622/51.  
 DR P-PSDB; AAB15200.  
 XX  
 PT New nucleic acid encoding a human diacylglycerol acyltransferase,  
 PT useful for treating hyperlipidaemia, atherosclerosis, heart disease, or  
 PT other diseases associated with an imbalance of triglyceride levels -

XX Claim 4; Fig 1B; 32pp; English.  
 XX  
 CC The enzyme acyl Coenzyme A-cholesterol acyltransferase 1 (ACAT1) mediates  
 CC sterol esterification, an important component of intracellular lipid  
 CC homeostasis. The present sequence is the coding sequence of human ACAT  
 CC Related gene Product 1 (ARGPI). The enzyme encoded by the present  
 CC sequence is a diacylglycerol acyltransferase (DGAT). ARGPI does not  
 CC esterify cholesterol. It is thought therefore that ARGPI participates in  
 CC the Coenzyme A-dependent acylation of substrate(s) other than cholesterol  
 CC e.g. diacylglycerol. Also, ARGPI has a predicted diacylglycerol binding  
 CC motif, suggesting that it may perform the last acylation in triglyceride  
 CC biosynthesis. ARGPI gene and protein are useful for treating a subject  
 CC who has an imbalance in triglyceride levels due to a defect in  
 CC esterification of diacylglycerol, via gene therapy. Particularly, ARGPI is  
 CC useful for treating hypertriglyceridaemia, hyperlipidaemia,  
 CC atherosclerosis, heart disease, obesity or other diseases associated with  
 CC high or excessive levels of triglyceride.  
 XX Sequence 1976 BP; 323 A; 539 C; 616 G; 398 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1.53e-65 Length: 1976  
 Score: 794.00 Matches: 205  
 Percent Similarity: 52.53% Conservative: 86  
 Best Local Similarity: 37.00% Mismatches: 190  
 Query Match: 28.65% Indels: 73  
 DB: 21 Gaps: 20  
 US-09-623-514A-2 (1-520) x AAA76169 (1-1976)  
 QY 4 LeuAspSerAlaGly-----ValThrThrValThrGluAsnGlyGlyGluPheVal 21  
 Db 91 GTTGACGCGCTGTGTAGCGGCTGACCGGCTACGCGCGCGGCGGCGGCGGCGGCGGCGG 150  
 QY 22 AspLeuAspArgLeuArgArg-----ArgLysSerArgSerAspSerSer 36  
 Db 151 CCGTTGTCTAGGCGCGGAGGTGGGCGCGGCGCTCGGCGGTACGAACCGCGAGGCC 210  
 QY 37 AsnGlyLeuLeuLeuSerGlySer----- 44  
 Db 211 CACGCTTGGCTGCGCGCGGCTGCGGCTGAGCCATGCGCGCGCGCGCGCGGCGGCGGCGG 270  
 QY 45 -----Asp-AsnAsnSerProSerAsp----AspValGlyAlaProAlaAspValArgAs 61  
 Db 271 CCGGAGGACAGGTGCGCGGCTCGAGCCACGCGCGGCGGCGGCGGCGGCGGCGGCGGA 330  
 QY 61 pArgIleAspSerValValAsnAspAspAlaGlnGly---ThrAlaAsnLeuAlaGlyAs 80  
 Db 331 AGAG-----GTGCGGAGCGCGCTGCGGCGCGCGCGCGCGCGCGCGCGGCGGCGG 378  
 QY 80 pAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGluGlyArgGlyAs 100  
 Db 379 CGCGCCAGCGG 438  
 QY 100 nAlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSe 120  
 Db 439 GCTGAGGTGC-----CATGCGCTGCAGGATTC 465  
 QY 120 rProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVa 140  
 Db 466 TTTATTTCAGCTCTGACAGCATGCCCGGTATTCTTCTGAGAACCTCATCAAGTATGTCAT 522  
 QY 140 lValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTr 160  
 Db 523 GGTGATCTCATCTTGGACATGCCCGGTATTCTTCTGAGAACCTCATCAAGTATGTCAT 582  
 QY 160 pLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAsp-----TrpPr 177  
 Db 583 CTTGGTG---GACCCCATCCAGGTGGTTCTCTGTTCTTCTGAAGGATCCCATAGCTGCC 639  
 QY 177 oLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLy 197

```

Db 361 AACTCTGGAGCTAAACATACCTTCAAGATGGCTCCGACACTCTGTTATCAAGCCAAAGT 420
Qy 300 TyrPro-ArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuVa 319
Db 421 TATCTTCGACACACCTTATGTAGAAAGGTGGTGGTCCGTCAGGTATATCTATATCT 480
Qy 319 lilePheThrGlyPheMetGlyPheIleleuGlnTyrIleAsnProIleValArgAs 339
Db 481 CATATTTACTGCTCCCAAGGATTCATTATTGAGCAATACATAAATCCTATTGTTGTGA 540
Qy 339 nSerLysHisProLeuLysGlyAspLeuLysGlyAspLeuLysGlyValLeuLysLeuSe 359
Db 541 CTCTCAACATCCATTGATGGGAGGATTACTGAATCTGTAGAGACTGTTTGAAGCTCTC 600
Qy 359 rValProAsnLeuTyrValTyrLeuLysMetPheTyrCysPhePheHisLeuTyrLeuAs 379
Db 601 ATTACCAATGCTACCTGCTGGCTTGGATGTTTATGCGCTTTCATCTGTGTTTAA 660
Qy 379 nileLeuAlaGluLeuLysCysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAl 399
Db 661 CATACTTCTGAGATCTTCGATTGTTGGTGACGAGAAATCTACAAAGACTGGTGAATGC 720
Qy 399 alySerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValar 419
Db 721 AAAGCAATATTGATGAGTACTGGAGAAATGGAACATGCTGTGCAATAAATGGATGTTTCG 780
Qy 419 gHistIleTyrPheProCysLeuArgSerLysLysProLysThrLeuAlaIleIleAl 439
Db 781 TCATATATATTCCTTCGATCGGAAATGGTATATCAAGGAAGTGTGCTGTTTTATATC 840
Qy 439 aPheLeuValSerAlaValPheHisGluLeuLysCysIleAlaValProCysArgLeuPheLy 459
Db 841 GTTCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900
Qy 459 sleuTrpIlePheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrIle 479
Db 901 GTTCTGGGCTTCTTAGGAATCATGCTTCAGATTCCTCCCTCATCATATTGACATCATACCT 960
Qy 479 uGlnGluArgPheGlySerThr---ValGlyAsnMetIlePheTrpPheIlePheCysTrl 498
Db 961 CAAATAAATATCATGACACAAATGGTGGCAATATGATCTTTGGTGTGTTTTCTGCTGAT 1020
Qy 498 ePheGlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArg 515
Db 1021 ATACGGCAGCAATGTGTTCTATTGTTATTTACCATGATGTGATGAACCGG 1072

RESULT 14
AAA48936
ID AAA48936 standard; cDNA; 901 BP.
XX
AC AAA48936;
XX
DT 06-DEC-2000 (first entry)
DE
DE Corn diacylglycerol acyltransferase cDNA #4.
XX
KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS Zea mays.
FX
FX Key Location/Qualifiers
FX CDS 1..858
FX /tag= a
FX /partial
FX /transl_except= (pos:442..444,aa:Xaa)
FX /transl_except= (pos:490..492,aa:Xaa)
FX /transl_except= (pos:587..589,aa:Xaa)
FX /transl_except= (pos:805..807,aa:Xaa)
FX /transl_except= (pos:820..822,aa:Xaa)
FX /note= "Xaa= unknown"
FX /product= Partial_diacylglycerol_acyltransferase

```

```

XX WO200032756-A2.
PN 08-JUN-2000.
XX
PD 01-DEC-1999; 99WO-US28354.
XX
PF 02-DEC-1998; 98US-0110602.
XX
PR 31-MAR-1999; 99US-0127111.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX
PI Cahoon EB, Kinney AJ, Cahoon RE;
XX
DR WPI; 2000-412308/35.
XX
DR P-PSDB; AAY94516.
XX
PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT synthesis of triacylglycerols and increasing the level of oils in plant
PT seeds -
XX
PS Claim 18; Page 46; 62pp; English.
XX
CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC soybean and wheat were screened for sequences with homology to a
CC putative acyl CoA cholesterol acyltransferase related gene from
CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is corn diacylglycerol acyltransferase
CC cDNA from a contig of clones p0042.cspaf49r, p0122.ckm57r and
CC p0125.czaa61rb. Diacylglycerol acyltransferases are involved in the
CC synthesis of triacylglycerols. Alteration of the expression of the
CC diacylglycerol acyltransferase DNA can be useful for increasing the
CC level of oils in plant seeds. Inhibitors of diacylglycerol
CC acyltransferase may be useful as herbicides.
XX
SQ Sequence 901 BP; 256 A; 172 C; 193 G; 276 T; 4 other;

```

```

Alignment Scores:
Pred. No.: 2,14e-88 Length: 901
Score: 1028.50 Matches: 188
Percent Similarity: 79.06% Conservative: 31
Best Local Similarity: 67.87% Mismatches: 48
Query Match: 37.12% Indels: 10
DB: 21 Gaps: 3

US-09-623-514A-2 (1-520) x AAA48936 (1-901)
Qy 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLys--- 271
Db 31 GTCCTTATGACATACAAATATATGATATAGGGTATGTCCTCAAAAGTACTGAGAGGCT 90
Qy 272 -----AlaAsnProGluValSerTyrTyrValSerLeuLysSerLeu 285
Db 91 GCTGCATATGGAATATTATGTCGATCTGAGATATGAAAGATCAACCTTTTAAAGTCTFA 150
Qy 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305
Db 151 GTGTACTTCATGTTGGCCCCAACACTTTGTTACCAGCAACTATTCCTCAAACTACATGT 210
Qy 306 IleArgLysGlyTyrPheValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325
Db 211 ATTAGAAAGGTTGGGTGACCCAGCACTCATAAAGTGGGTGTTTACAGCGCTTGATG 270
Qy 326 GlyPheIleleuGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345
Db 271 GGCCTTCAATATGACCAATATATAAACCCTGTTGAGAGAAATTCCTCAAACTCACTGAAA 330
Qy 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365
Db 331 GGGAAATTTTGAATGCTATAGAAAGAGTCTTTAAACACTCTCAGTGCCCAACATTATATGTA 390

```

Db 601 TCATTACCAAAATGCTACCTGGCTTTCATGCTTTTATTCCTTTTCCATCTGTGGTGA 660  
 QY 379 AsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpAsn 398  
 Db 661 AACATACCTGCTGAGATCTTCGATTGGTGGAGAGAAATCTACAAAGACTGGTGAAT 720  
 QY 399 AlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetVal 418  
 Db 721 GCAAGACAATGATGAGTACTGGAGAAATGGAGCATGCTGTGCATAAATGGATTGT 780  
 QY 419 ArgHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIle 438  
 Db 781 CGTCATATATATTTCTTCGATCGCAATGGTATATCAAGGAAGTCTGTTTATTA 840  
 QY 439 AlaPheLeuValSerAlaVal-Phe 446  
 Db 841 TCGTTCTTGTCTCTGCTACTTTCATGAGTAACATTATTTACTTTTTCACCTCTTCATCT 900  
 QY 446 446  
 Db 901 GCATATATTAAATATATAGTCTCTATTTTCAAAATGTCTTTTCGAGTTTCGACATGCT 960  
 QY 446 446  
 Db 961 TTGTTCAAACTTACCAGCTGTAGATTACTTGGATGAAGTCTCTATATAAAATTCATTA 1020  
 QY 447 447  
 Db 1021 TTTCAAAATCCAGTCCCTTTCCAGAAAATATGATACATTTTGTTCGATTTTGTACACCA 1080  
 QY 448 448  
 Db 448 uLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPh 468  
 Db 1081 GTTATCGGTGAGTTCCTCCCTCCACATATCAAGTCTGGCTTCTTAGGAATCATGCT 1140  
 QY 468 eGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThr 487  
 Db 1141 TCAGATTCCTTCATCATATATGATCATACCTCAAAATATAATTCAGTCACACATGGT 1200  
 QY 487 lGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLe 507  
 Db 1201 TGGCATATGATCTTTTGTGTTTTTTCGATATACGGCAGCCAATGTGTCTATT 1260  
 QY 507 uTyrTyrHisAspLeuMetAsnArg 515  
 Db 1261 GTATTACCAATGATGTGATGAACCGG 1285

## RESULT 13

AAA48933  
 ID AAA48933 standard; cDNA; 1281 BP.  
 XX  
 AC AAA48933;  
 XX  
 DT 06-DEC-2000 (first entry)  
 XX  
 DE Corn diacylglycerol acyltransferase cDNA #1.  
 XX  
 KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;  
 KW triacylglycerol; herbicide; EC2.3.1.20; ss.  
 XX  
 OS Zea mays.  
 XX  
 FH Key  
 XX Location/Qualifiers  
 FT CDS  
 FT 1..1090  
 FT /tag- a  
 FT /partial  
 FT /transl\_except= (pos:283..286,aa:Ile)  
 FT /note= "This codon has an apparent 1 nucleotide  
 FT insertion which alters the reading frame"  
 FT /transl\_except= (pos:356..358,aa:xaa)  
 FT /note= "xaa= unknown"  
 FT /transl\_except= (pos:869..889,aa:YVLLFL)  
 FT /product= Partial\_diacylglycerol\_acyltransferase  
 FT  
 XX

PN WO200032756-A2.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 01-DEC-1999; 99WO-US28354.  
 XX  
 PR 02-DEC-1998; 98US-0110602.  
 PR 31-MAR-1999; 99US-0127111.  
 XX  
 PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Cahoon EB, Kinney AJ, Cahoon RE;  
 XX  
 DR WPI; 2000-412308/35.  
 DR P-PSDB; AAY94513.  
 XX  
 PT Polynucleotides encoding diacylglycerol acyltransferase, useful for  
 PT synthesis of triacylglycerols and increasing the level of oils in plant  
 PT seeds  
 XX  
 PS Claim 15; Page 40-42; 62pp; English.  
 XX  
 CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,  
 CC soybean and wheat were screened for sequences with homology to a  
 CC putative acyl CoA cholesterol acyltransferase related gene from  
 CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo  
 CC sapiens and Mus musculus. The cDNA clones identified from this process  
 CC were used to form complete diacylglycerol acyltransferase cDNA  
 CC sequences. The present sequence is corn diacylglycerol acyltransferase  
 CC cDNA from a contig of clones cpjlc.pk005.k23, cen3n.pk0010.c10,  
 CC ccol.pk0029.b6. Diacylglycerol acyltransferases are involved in the  
 CC synthesis of triacylglycerols. Alteration of the expression of the  
 CC diacylglycerol acyltransferase DNA can be useful for increasing the  
 CC level of oils in plant seeds. Inhibitors of diacylglycerol  
 CC acyltransferase may be useful as herbicides.  
 XX  
 SQ Sequence 1281 BP; 351 A; 256 C; 245 G; 417 T; 12 other;

Alignment Scores:  
 Pred. No.: 2.98e-110 Length: 1281  
 Score: 1260.00 Matches: 229  
 Percent Similarity: 76.94% Conservative: 48  
 Best Local Similarity: 63.61% Mismatches: 71  
 Query Match: 45.47% Indels: 12  
 DB: 21 Gaps: 2

US-09-623-514A-2 (1-520) x AAA48933 (1-1281)

QY 168 PheSerSerArgSerLeuArgAspTrpProLeuPheMetCysLysIleSerLeuSerIle 187  
 Db 1 TTTAATGCTACATCATTCGAGACTGGCCACTGCTAATGTGTGCTTAGCTACCCATA 60  
 QY 188 PheProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIleSerGluPro 207  
 Db 61 TTTCCCTTGGTGCATTTTCAGTCGAAAGTGGCATTCACAAATTCGTGTAGTATCCT 120  
 QY 208 ValValIlePheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrVal 227  
 Db 121 GCTACTACCTGTTTTCACATCTCTTTTACAAACATTTGAATTTGATATATATCAGTCTG 180  
 QY 228 ThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIle 247  
 Db 181 ATTTTAAAGTGTGATTTCTGCAGTTTATCAGGCTTTGTGTGTATGTTATTTGCTGCATT 240  
 QY 248 ValTrpLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg-SerLeuAlaAs 267  
 Db 241 GTTTGGCTGAAGCTTCTTCTTTTGGACATACAAACCATGATATACGAAATGATCAC 300  
 QY 267 nAlaAlaAspLysAlaAsnProGluValSer-----TyrTyrVa 280  
 Db 301 AAGCGGCAAGAGGTGATAATGAAGTACCGCGGCTGGCATAGATAAATTTACAANCTCC 360  
 QY 280 lSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGln-Pro-Ser 299





1445 AATAATGATTTTGGTTCATATTCAGTATCCCTTGGTCAACCTATGTGTACTGTCTATAC 1504  
 509 TyrHisAspLeuMetAsnArgLysGlySerMet 519  
 1505 TACCATGACTTGATCAATAGGAAGGCAACTT 1537

## RESULT 10

AAAA48938  
 ID AAA48938 standard; cDNA: 1587 BP.

XX AC AAA48938;

XX XX 06-DEC-2000 (first entry)

XX Rice diacylglycerol acyltransferase cDNA #2.

XX Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX triacylglycerol; herbicide; EC2.3.1.20; ss.

XX Oryza sativa.

XX Key Location/Qualifiers

XX CDS 15..1310

XX /tag= a

XX /partial

XX /note= "The CDS of this sequence only encodes

XX amino acids 70 to 500 of the protein in AAY94518"

XX PN WO200032756-A2.

XX PD 08-JUN-2000.

XX XX 01-DEC-1999; 99WO-US28354.

XX PR 02-DEC-1998; 98US-0110602.

XX PR 31-MAR-1999; 99US-0127111.

XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.

XX PI Cahoon EB, Kinney AJ, Cahoon RE;

XX WPI; 2000-412308/35.

XX DR P-PSDB; AAY94518.

XX Polynucleotides encoding diacylglycerol acyltransferase, useful for  
 XX synthesis of triacylglycerols and increasing the level of oils in plant  
 XX seeds -

XX Claim 4; Page 48-49; 62pp; English.

XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,  
 XX soybean and wheat were screened for sequences with homology to a  
 XX putative acyl CoA cholesterol acyltransferase related gene from  
 XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo  
 XX sapiens and Mus musculus. The cDNA clones identified from this process  
 XX were used to form complete diacylglycerol acyltransferase cDNA  
 XX sequences. The present sequence is rice diacylglycerol acyltransferase  
 XX cDNA from clone rls24.pk0034.d8.fis. Diacylglycerol acyltransferases are  
 XX involved in the synthesis of triacylglycerols. Alteration of the  
 XX expression of the diacylglycerol acyltransferase DNA can be useful for  
 XX increasing the level of oils in plant seeds. Inhibitors of  
 XX diacylglycerol acyltransferase may be useful as herbicides.

XX Sequence 1587 BP; 411 A; 323 C; 334 G; 519 T; 0 other;

## Alignment Scores:

Pred. No.:	2.65e-142	Length:	1587
Score:	1597.50	Matches:	284
Percent Similarity:	80.05%	Conservative:	61
Best Local Similarity:	65.89%	Mismatches:	73
Query Match:	57.65%	Indels:	13
DB:	21	Gaps:	3

US-09-623-514a-2 (1-520) x AAA48938 (1-1587)

QY	93	GlyGlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSerValPro	112
DB	15	GGCGTGGGAC-----TTCTCCGGTTCACGTTCCCGCGCGCGCGCG	59
QY	113	AlaHisArgAlaArgGluSerProLeuSerSerAspAlaIlePheLysGlnSerHis	132
DB	60	GTGCACCGCAAGGCCAAGGAGAGCCCTCAGCTCCGAGCCCATCTTCAAGCAGAGTCAT	119
QY	133	AlaGlyLeuPheAsnLeuCysValValValLeuLeuAlaValAsnSerArgLeuIleIle	152
DB	120	GCAGGCTTTTCAACCTATGCAATGTTGTCTAGTTGCAGTGAACGAGGAGGCTTATTATC	179
QY	153	GluAsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSer	172
DB	180	GAGAACTTAAGAAGTATGGCTTATTAAGAAGCTGGGTTTGGTTTATGATAAATCA	239
QY	173	LeuArgAspTrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAla	192
DB	240	TTGCGGAGCTGGCCACTTCTAATGTGTCTTAGTCTGCCGTCTTCCCTCGGTGCA	299
QY	193	PheThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValIlePheLeu	212
DB	300	TTTGCAAGTTGAAAGTTGGCATTTAACCAATGTTATTACTGATGCTGTTGCTACCTGCCTC	359
QY	213	HisIleIleIleThrMetGluValLeuTyrProValTyrValThrLeuArgCysAsp	232
DB	360	CATATCTTCCTTTCAACCAACCGAAATGTATATCCAGTCTGTGATGCTTAAAGTGTGAT	419
QY	233	SerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeu	252
DB	420	TCTGCAGTTTGTCTGGCTTTTGTGATATTTATGCTGTATTGTTGGTGAAGCTT	479
QY	253	ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspIleAla	272
DB	480	GTATCTTTTGCACATACAAACCATGATATAGGCAACTGACCATGGCGGCAAGAAGTT	539
QY	273	AsnProGluValSerTyrTyr-----ValSerLeuLysSerLeu	285
DB	540	GATAATGAACCTAAGCACAGTTGCATGGATATTTACAACTCCCACTTTAGGAAATCTTA	599
QY	286	AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys	305
DB	600	ATATACTTCATGATGGCTCCCTACACTCTGTATACAGCAAGCATATCCCGAAGTTCATGT	659
QY	306	IleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet	325
DB	660	GTTAGAAAAGGTGGCTGATTCGTCAAATTTATCTGTACTTGTGATCTTACTGCTTCAA	719
QY	326	GlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys	345
DB	720	GGCTTCATATTGACCAATACATAATCAATGTTGTGAATTCACCATCCATCCATGAA	779
QY	346	GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal	365
DB	780	GGAGGACTCTAAATGCTGTAGAGACTGTTTGAACCTCTCATTTACCAATGTTTACCTG	839
QY	366	TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeu	385
DB	840	TGGCTTTCATGTTCTATGCTTTTCCATCTCTGTTTAAAGTATATCTTCTGCTGAGATCTT	899
QY	386	CysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyr	405
DB	900	CGATTTGGTGACCGGAAATCTTACAAAGATTGGTGAATGCAAAACCAATTCATGATAT	959
QY	406	TrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCys	425
DB	960	TGGAGAAATGGAATATGCTGTACATAAATGGGTGTGTCGCATATTTACTTTCTTCTGC	1019
QY	426	LeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaVal	445



```

|||||
3081 TTTGATGCTCTCACTTGCATTTGCTGGCTAAAGTTGGTTCTTATGCTCATACATAGCTA 3140
QY
260 rAspileArgSerLeuAlaAsnAlaAlaAspLysAla----- 272
Db
3141 TGACATAAGATCCCTAGCCAAATGCAGCTGATAGCT-AAAAACGAAAAAAGACGCTATG 3199
QY
273 -----As 273
Db
3200 TATTAGTCACCTTGCACCTGCTTTACTGTTTAAACCAACACACTGTTATGAACCTTTAGGCCAA 3259
QY
273 pProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetValAlaProTh 293
Db
3260 TCCTGAAGTCTCCTACTAGCTTGAAGAGCTTGGCATATTTTCATGGTGCCTCCAC 3319
QY
293 rLeuCysTyr----- 296
Db
3320 ATTGTGTTATCAGGTAAGTGCAGCAAGTCATCAACCATCTCTATATCTTGCAGAGTTTCTT 3379
QY
297 -----GlnProSerTyrProArgSerAlaCysIleAr 307
Db
3380 GTCTAAACCTCGGATCTTTGCTTTTCCCGAGCAAGTTATCCACGTTCTGCTATATACG 3439
QY
307 gLysGlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPh 327
Db
3440 GAAGGGTTGGGGGCTGCTCAATTTGCAGAACTGGTCATATTTACCGGATTCATGGGATT 3499
QY
327 eileIleGlu----- 330
Db
3500 TATATAGAACAAAGTACGTTTTCACATCTTGTCTTATTAGTTTTCCTTGGTGAATCAT 3559
QY
331 -----GlnTyrIleA 334
Db
3560 CATCCCTGGTGTGCACCATGACTTCATGTTCTTTGTTACATTTTGGCAGTATATAA 3619
QY
334 snProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluA 354
Db
3620 ATCCCTATTGTAGGAACCAAGCATCCTTTGAAAGGCGATCTTCTATATGCTATNTGAA 3679
QY
354 rGValLeuLysLeuSerValProAsnLeuTyrValTrpLeuCysMetPheTyrCysPheP 374
Db
3680 GAGTCTTGAGCTTTTCAGTCCCAATTTATATGTGTGGCTCGCATGTTCTACTGCTTCT 3739
QY
374 heHisLeuTrp----- 377
Db
3740 TCCACCTTTGTGTATGCTGTGATCCCATCTCTTTCAAATAAATTCGAAAAACC 3799
QY
377 ----- 377
Db
3800 GAAAAAGGCTAAATCTCATACGAATTTGATATTTTATGTTTCTTAGAGTCGGTGATGAA 3859
QY
378 -----LeuAsnIleLeuAlaGluLeuLeuC 386
Db
3860 TTTTCAGTTACTGAACGCAAAATCTCTTGCCAAAGGTTAAACATATTTGGCAGAGCTTCT 3919
QY
386 ysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGlyAsp---- 404
Db
3920 GCTTCGGGGATGCTGAATTTCTACAAAGATTTGGTGAATGCAAAAAGTGTGGGAGA-TGTG 3978
QY
404 ----- 404
Db
3979 ACCTATTTTACTCAAAGAAAACTATGATTTTAAATGTTGTCGGTGTGTTTGGGTGTCATC 4038
QY
405 -----TyrTrpArgMetTrpAsnM 411
Db
4039 TAACTAACCAAAATTCATGATTCATCTGCTTCCTTTATCATGACTGAGTGGAGATGTGGAATA 4098
QY
411 et----- 411
Db
4099 TGGTATGTTCTCTCCCTAAACATCACCTTCTTTGTACACAAAAATAGAAGAAGAGAGCT 4158
QY
412 -----ProValHisLysTrpMetValArgHisIleTyrP 423
|||||

```

---

```

4159 AATTAAAGATCTTGTGTTTCTCTTGACAGCCCTGTTTCATATAAATGGATGGTTCGACATATATACT 4218
QY
423 heProCysLeuArgSerLysIleProLys----- 432
Db
4219 TCCCGTGTGTTGGCAGCAAGATACCAAA-GGTGAGTGAGATATATACCGATATGCAATTTG 4277
QY
433 -----ThrLeuA 435
|||||
4278 TCGAGATTTGTTTCTGTGATATAAATTTAACCCCTCCACACACTTGTGTTTTTTCAGACACTCG 4337
QY
435 laileIleIleAlaPheLeuValSerAlaValPheHisGlu----- 448
|||||
4338 CCATATATCAATGCTTTTCTTAGTCTCTGCAGTCTTTTCATGAGGTATACATACATCTTCTACAT 4397
QY
448 ----- 448
Db
4398 TCCCTGCTCTCTAGAGCATGAACACACACCTAGTGAAGAAATGCTAAATATTCAAAGCAT 4457
QY
449 -----LeuCysIleAlaValP 454
|||||
4458 TGTGTTTTACTTAACGATCTTGTGTTACAAAATTTCTTTTGACAGCTATGATCGCAGATTTC 4517
QY
454 roCysArgLeuPheLysLeuTTPAlaPheLeuGlyIleMetPheGln----- 469
Db
4518 CTTGTGCTCTCTCAAGCTATGGGCTTTTCTTGGGATTAATGTTTCAGGTTAAAAAATTAC 4577
QY
469 ----- 469
Db
4578 TAAACTGCTGCAGTCGATTTTACTTAAACTCTAATCTCATATTTCTCACCACCAACCAATTTGT 4637
QY
470 -----ValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThrV 487
|||||
4638 TTGAGTAGTGCTCTTTGGTCTTCATCACAAACTATCTACAGGAAGGTTTGGCTCAACGG 4697
QY
487 al----- 487
Db
4698 T-ATGCTCTCAAAACCCGAGAAAAATAGAACGAATAACTCTTCTTTCATAGCTAGCCAT 4756
QY
487 ----- 487
Db
4757 TTAAATCGCAATGCTGAAACTTAAATAAAGGTGATCTGTTTGAATGGGATCATATT 4816
QY
488 -----GlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnPrometCysV 505
|||||
4817 ATTAGTGGGGAACATGATCTTCTGTTTCATCTTCTGCAATTTTCGACAAACCGATGTGTG 4876
QY
505 alLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
|||||
4877 TGCTTCTTTATTACCACGACCTGATGACCCGAAAGGATCGATGTCA 4923
Db

```

RESULT 9  
AAA48939  
ID AAA48939 standard; cDNA; 1942 BP.  
XX  
AC AAA48939;  
XX  
DT 06-DEC-2000 (first entry)  
XX  
DE Soybean diacylglycerol acyltransferase cDNA #1.  
XX  
KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;  
KW triacylglycerol; herbicide; EC2.3.1.20; ss.  
XX  
OS Glycine max.  
XX  
FH Key Location/Qualifiers  
FT CDS 29..1543  
FT /\*tag- a  
FT /product= Diacylglycerol\_acyltransferase  
XX  
PN WO200032756-A2.  
XX  
PD 08-JUN-2000.



FT	exon	/tag- af 4646..4696
FT	Intron	/tag- ag 4697..4821
FT	exon	/tag- ah 4822..4926
FT	FT	/tag- ai
FT	FT	/partial
XX	WO200036114-A1.	
PN	XX	
PD	22-JUN-2000.	
XX	XX	
PF	16-DEC-1999;	99WO-CA01202.
PR	17-DEC-1998;	98US-O112812.
XX	(CANADA ) NAT RES COUNCIL CANADA.	
PA	Zou J, Taylor DC, Wei Y, Jako CC;	
PI	WPI; 2000-431592/37.	
XX	P-PSDB; AAY96854.	
DR	New DNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana for transforming plants and regulating seed oil content, fatty acid synthesis and seed oil acyl composition in commercial and crop plants	
PT	Disclosure; Page 79-81; 91pp; English.	
XX	XX	
CC	This DNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11 diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a 147 bp insertion located at the central region of intron 2. The insertion is a duplication of a segment that is composed of 12 bp from the 3' end of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful for regulating seed oil content, the ratio of diacylglycerol to triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil acyl composition, seed size/weight and carbon flux into other seed components in commercial and crop plants. The natural formation of triacylglycerols can be modified to increase the yield in commercial plant oils or modify their composition to achieve specific commercial improvements of plants and plant products.	
CC	Sequence 5339 BP; 1620 A; 919 C; 930 G; 1870 T; 0 other;	
XX	XX	
SQ	Alignment Scores:	
Pred. No.:	1.43e-200	Length: 5339
Score:	2216.50	Matches: 515
Percent Similarity:	48.72%	Conservative: 0
Best Local Similarity:	48.72%	Mismatches: 5
Query Match:	79.99%	Indels: 541
DB:	21	Gaps: 15
US-09-623-514A-2 (1-520) x AAA51485 (1-5339)		
QY	1 MetAlaIleLeuAspSerAlaGlyValThrValThrValThrGluAsnGlyGlyGluPhe 20	
Dd	1761 ATGGCGATTTCGTCTGCTGGCTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTTC 1820	
QY	21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40	
Dd	1821 GTCGATCTTGATAGGCTTCGTGACGAGAAATCGAGATCGGATCTTCTTAACGGACTTCTT 1880	
QY	41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60	
Dd	1881 CTCCTGGTTCGGATAAATCTCCTTCGGATGATGTGGAGCTCCGCCGAGCTTAGG 1940	
QY	61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyasp 80	
Dd	1941 GATCGGATTGATCCGTTGTTNACCATGACGCTCAGGNACAGCCAATTTGGCCGGAGAT 2000	

```
QY 404 ----- 404
Db 3859 TGAATTTTAAAGTGTGCGTGTGTTTGGGTCTACTAACTAACAAATTCATGTATTCACT 3918
QY 405 -----TyrTrpArgMetTrpAsnMet----- 411
Db 3919 GTCTTCCTTATCAGTACTGGAGATGTCGAATATGGTATGTTCTCTCTCCCTTAACATCA 3978
QY 411 ----- 411
Db 3979 CCTTCCTTTGTACACAAATAGAGAAGAGAGCTAATTAAGATCTTTGTTTCTCCTGCACAG 4038
QY 412 ProValHisLysTrpMetValargHisIleTyrPheProCysLeuArgSerLysIlePro 431
Db 4039 CCTGTTCATAAATGGATGTTCCGACATATATCTCCCGTGTGTCGACAGATACCA 4098
QY 432 Lys----- 432
Db 4099 AA-GGTGAGTGAGATATATACCGATATGCAATGTGCGAGATTTGTTCTGTGATATAAAT 4157
QY 433 -----ThrLeuAlaIleIleIleAlaPheLeuValser 443
Db 4158 TTAACCCCTCCACACACTTGTGTTTTCAGACACTCGCCATTATCATTCGTTCTCCTAGTCTCT 4217
QY 444 AlaValPheHisGlu----- 448
Db 4218 GCAGTCTTTTCATGAGGTATACATACATCTTCTACATTCGCCCTGTCTCTAGACGCATGAACAC 4277
QY 448 ----- 448
Db 4278 ACGTAGTGAAGAAATGCTAATATCAAGCATGTTGTTTACTTAACGATCTGTGTGTTA 4337
QY 449 -----LeuCysIleAlaValProCysArgLeuPheLysLeuTrpAla 462
Db 4338 CAAATTTCCCTTTTGACAGCTATGCATCGCAGTTCCTTGTCTCTCTCAAGCTATGGCT 4397
QY 463 PheLeuGlyIleMetPheGln----- 469
Db 4398 TTCTTGGGATTATGTTTTCAGGTTAAATAAATTACTAACTGCTGCAGTCGATTTTACTA 4457
QY 470 -----ValProLeuValPheIle 475
Db 4458 AACTCTAATCTCATATTCGACCAACCAATTTGTTGAGTAGTGCTTTGGTCTTCATC 4517
QY 476 ThrAsnTyrLeuGlnGluArgPheGlySerThrVal----- 487
Db 4518 ACAAACTATCTACAGGAAGGTTTGCTCAACGGT-ATGCTCTCAAAACCCGAGAAATA 4576
QY 487 ----- 487
Db 4577 GAACGAATAACTCTTTTCTTTCATAGCTAGCCATTTAAATCGCAATGCTGAAACTTAATA 4636
QY 488 -----GlyAsnMetIlePheTrp 493
Db 4637 ATAAAGGTGATCTGTTTGGATGGATGATATATATAGGTGGGGAACATGATCTCTCG 4696
QY 494 PheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMet 513
Db 4697 TTCATCTTCTGCATTTTCGACAAACCGATGTGTGCTCTCTTTATTACCAACCGCTGATG 4756
QY 514 AsnArgLysGlySerMetSer 520
Db 4757 AACCGAAAAGGATCGATGTCA 4777

RESULT 8
AAA51485
ID AAA51485 standard; DNA; 5339 BP.
XX
AC AAA51485;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana AS11 diacylglycerol acyltransferase genomic DNA.
```

```
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
OS size; weight; carbon flux; TAG1; insertion mutant; ds.
XX Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1761..4926
FT exon /tag= a
FT exon 1761..2150
FT /tag= b
FT /partial
FT Intron 2151..2234
FT /tag= c
FT exon 2235..2315
FT /tag= e
FT Intron 2316..2380
FT /tag= f
FT misc_RNA 2370..2515
FT /tag= d
FT exon /note= "insertion region"
FT 2381..2461
FT /tag= g
FT /note= "duplication of exon 2"
FT Intron 2462..2552
FT /tag= h
FT exon 2553..2622
FT /tag= i
FT Intron 2623..2728
FT /tag= j
FT exon 2729..2804
FT /tag= k
FT Intron 2805..2901
FT /tag= l
FT exon 2902..2970
FT /tag= m
FT Intron 2971..3050
FT /tag= n
FT exon 3051..3174
FT /tag= o
FT Intron 3175..3234
FT /tag= p
FT exon 3235..3332
FT /tag= q
FT Intron 3333..3410
FT /tag= r
FT exon 3411..3512
FT /tag= s
FT Intron 3513..3612
FT /tag= t
FT exon 3613..3749
FT /tag= u
FT Intron 3750..3893
FT /tag= v
FT exon 3894..3975
FT /tag= w
FT Intron 3976..4079
FT /tag= x
FT exon 4080..4100
FT /tag= y
FT Intron 4101..4184
FT /tag= z
FT exon 4185..4247
FT /tag= aa
FT Intron 4248..4330
FT /tag= ab
FT exon 4331..4378
FT /tag= ac
FT Intron 4379..4501
FT /tag= ad
FT exon 4502..4564
FT /tag= ae
FT Intron 4565..4645
```



```

Db 1459 CATAAATGGATGGTTGACATATATATCTTCCCGGCTTGGCAGCAAGATACCAAGACA 1518
Qy 434 LeuAlaIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAlaVal 453
Db 1519 CTCGCCATTATCATGCTTCTCTAGTCTCTGCACTCTTCATGAGCTATGATCGCAGTT 1578
Qy 454 ProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuVal 473
Db 1579 CCTTGTGCTCTCTCAAGCTATGGCTTTCTTGGGATTATGTTTCAGGTGCCCTTGGTC 1638
Qy 474 PheIleThrAnTyLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePheTrp 493
Db 1639 TTCATCAAAACTATACAGGAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCGG 1698
Qy 494 PheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyTrpHisAspLeuMet 513
Db 1699 TTCATCTCTGCATTTTCGGACACCGATGTGTGCTCTCTTTATACCAGGACCTGATG 1758
Qy 514 AsnArgLysGlySerMetSer 520
Db 1759 AACCGAAAGGATCGATGCA 1779

RESULT 7
AAAS1483
ID AAA51483 standard; DNA; 5193 BP.
AC
XX
XX
XX 09-OCT-2000 (first entry)
XX
XX
XX A. thaliana diacylglycerol acyltransferase genomic DNA.
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; ds.
XX Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1761..4780
FT FT /*tag= a
FT exon 1761..2150
FT FT /*tag= b
FT FT /partial
FT intron 2151..2234
FT FT /*tag= c
FT FT 2223..2369
FT FT /*tag= d
FT FT /note= "region duplicated in insertion mutant AS11
FT FT (see AAA51485)"
FT FT 2235..2315
FT FT /*tag= e
FT intron 2316..2406
FT FT /*tag= f
FT exon 2407..2480
FT FT /*tag= g
FT intron 2481..2582
FT FT /*tag= h
FT FT 2583..2661
FT FT /*tag= i
FT intron 2662..2758
FT FT /*tag= j
FT exon 2759..2823
FT FT /*tag= k
FT intron 2824..2927
FT FT /*tag= l
FT exon 2928..3028
FT FT /*tag= m
FT intron 3029..3108
FT FT /*tag= n
FT exon 3109..3186
FT FT /*tag= o
FT intron 3187..3264
FT FT /*tag= p

```

```

FT exon 3265..3366
FT FT /*tag= q
FT intron 3367..3466
FT FT /*tag= r
FT exon 3467..3603
FT FT /*tag= s
FT intron 3604..3747
FT FT /*tag= t
FT exon 3748..3829
FT FT /*tag= u
FT intron 3830..3933
FT FT /*tag= v
FT exon 3934..3954
FT FT /*tag= w
FT intron 3955..4038
FT FT /*tag= x
FT exon 4039..4101
FT FT /*tag= y
FT intron 4102..4184
FT FT /*tag= z
FT exon 4185..4232
FT FT /*tag= aa
FT intron 4233..4355
FT FT /*tag= ab
FT exon 4356..4418
FT FT /*tag= ac
FT intron 4419..4499
FT FT /*tag= ad
FT exon 4500..4550
FT FT /*tag= ae
FT intron 4551..4675
FT FT /*tag= af
FT exon 4676..4780
FT FT /*tag= ag
FT FT /partial
XX
XX WO200036114-A1.
XX
XX 22-JUN-2000.
XX
XX 16-DEC-1999; 99WO-CA01202.
XX
XX 17-DEC-1998; 98US-0112812.
XX (CANA ) NAT RES COUNCIL CANADA.
XX
XX Zou J, Taylor DC, Wei Y, Jako CC;
XX
XX WPI; 2000-431592/37.
XX P-PSDB; AAY96853.
XX
XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis
XX thaliana for transforming plants and regulating seed oil content, fatty
XX acid synthesis and seed oil acyl composition in commercial and crop
XX plants
XX
XX Claim 2; Page 67-68; 91pp; English.
XX
XX This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
XX (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
XX seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
XX in seed oil, fatty acid synthesis, seed oil acyl composition, seed
XX size/weight and carbon flux into other seed components in commercial and
XX crop plants. The natural formation of triacylglycerols can be modified to
XX increase the yield in commercial plant oils or modify their composition
XX to achieve specific commercial improvements of plants and plant products.
XX
XX Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;
SQ

```

```

Alignment Scores:
Pred. No.: 2.12e-201 Length: 5193
Score: 2225.00 Matches: 515
Percent Similarity: 51.04% Conservative: 0

```

CDS 139..1782  
 FT /tag- a  
 FT /product- diacylglycerol\_acyltransferase  
 FT /note- "Insertion mutant"  
 FT misc\_RNA 610..690  
 FT /tag- b  
 FT /note- "81 bp insertion, duplication of exon 2"  
 XX  
 PN WO200036114-A1.  
 XX  
 XX 22-JUN-2000.  
 XX  
 XX 16-DEC-1999; 99WO-CA01202.  
 XX  
 XX 17-DEC-1998; 98US-0112812.  
 XX  
 XX (CANADA) NAT RES COUNCIL CANADA.  
 XX  
 XX Zou J, Taylor DC, Wei Y, Jako CC;  
 XX  
 XX WPI; 2000-431592/37.  
 XX P-PSDB; AAY96854.  
 XX  
 XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis  
 PT thaliana for transforming plants and regulating seed oil content, fatty  
 PT acid synthesis and seed oil acyl composition in commercial and crop  
 PT plants  
 XX  
 XX Claim 5; Page 79; 9lpp; English.  
 XX  
 XX This cDNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11  
 CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a  
 CC 147 bp insertion located at the central region of intron 2. The insertion  
 CC is a duplication of a segment that is composed of 12 bp from the 3' end  
 CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the  
 CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful  
 CC for regulating seed oil content, the ratio of diacylglycerol to  
 CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil  
 CC acyl composition, seed size/weight and carbon flux into other seed  
 CC components in commercial and crop plants. The natural formation of  
 CC triacylglycerols can be modified to increase the yield in commercial  
 CC plant oils or modify their composition to achieve specific commercial  
 CC improvements of plants and plant products.  
 XX  
 XX Sequence 1985 BP; 482 A; 415 C; 446 G; 642 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 7,89e-252 Length: 1985  
 Score: 2747.50 Matches: 520  
 Percent Similarity: 95.06% Conservative: 0  
 Best Local Similarity: 95.06% Mismatches: 0  
 Query Match: 99.15% Indels: 27  
 DB: 21 Gaps: 1  
 US-09-623-514A-2 (1-520) x AAA51484 (1-1985)  
 QY 1 MetAlaIleuAspSerAlaGlyValThrThrValThrCluAsnGlyGlyGluPhe 20  
 DB 139 ATGGCGATTGTGATTCTCTGCGGTTACTACGGTGACGAGAACGGTGGCGGAGAGTTC 198  
 QY 21 ValAspLeuAspArgLeuArgArgGlySerArgSerAspSerSerAsnGlyLeuLeu 40  
 DB 199 GTGATCTTGATAGGCTTCGTCGACGGAATCGATCGGATCTTCTACGACATCTCT 258  
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
 DB 259 CTCCTCTGGTCCGATAATAATCTCTCGGATGATGTTGGAGCTCCCGCGGACGTTAGG 318  
 QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
 DB 319 GATCGGATGATTCCTGTTTACGATGACGCTCAGGGAACAGCCCAATTTGGCCGGAGAT 378  
 QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGluGlyArgGlyAsn 100

DB 379 AATAACGGTGGTGGCGATAATAACGGTGGTGGAGAGCGCGGAGAGGAAGAAAC 438  
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120  
 DB 439 GCCGATGCTAGCTTACGATATCGACCGTGGTTCACGCTCATCGGAGGCGAGAGAGT 498  
 QY 121 ProLeuSerSerAspAlaIlePheLysGln----- 130  
 DB 499 CCACCTTAGCTCCGACGCAATCTTCAACAGAGCCATGCCGGATTATTCAACCTCTGTGTA 558  
 QY 131 -----SerHisAla 133  
 DB 559 GTAGTTCTTATTGCTGTAAACACGTAGACTCATCATCGAAAAATCTTATGAAGAGCCATGCC 618  
 QY 134 GlyLeuPheAsnLeuCysValValValLeuIleAlaValAsnSerArgLeuIleGlu 153  
 DB 619 GGATTATTCAACCTCTGTGTAGTCTTATTGCTGTAAACAGTACATCATCATGAA 678  
 QY 154 AsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSerLeu 173  
 DB 679 AATCTTATGAAGTATGGTTGGTTGATCAGAACGATTTCTGTTTAGTTCAAGATCGCTG 738  
 QY 174 ArgAspTrpProLeuPheMetCysCysIleSerLeuSerIlePhePheProLeuAlaIlePhe 193  
 DB 739 CGAGATTGGCGCTTTTCATGTGTGTATATCCCTTTCGATCTTCTCTTTGGGTGCCCTTT 798  
 QY 194 ThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeuHis 213  
 DB 799 ACGTTTGAGAAATGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTCTTCAT 858  
 QY 214 IleIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSer 233  
 DB 859 ATTATTATCACCATGACAGAGGTTTGTATCCAGTTTACGTACCCCTTAAGGTGTGATCT 918  
 QY 234 AlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuVal 253  
 DB 919 GCATTTTATCAGGTGTCACTTTGATGCTCCTCCTCCTGTCATTTGTGGCTTAAAGTTGGTT 978  
 QY 254 SerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsn 273  
 DB 979 TCTTATGCTCATACTAGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 1038  
 QY 274 ProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetValAlaProThr 293  
 DB 1039 CCTGAAGTCTCTCTACTAGTTAGCTTGAAGAGCTTGGCATATTTTCATGCTGCTCCACA 1098  
 QY 294 LeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTrpValAlaArg 313  
 DB 1099 TTGTGTTATCAGCAAGTATATCCAGTTCGTCATGATATACGGAAGGTTGGTGGCTCGT 1158  
 QY 314 GlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleLeuGlnTyrIle 333  
 DB 1159 CAATTTGCAAACTGGTTCATATTCACCGGATTCACCGGATTCATATAGAACATATATA 1218  
 QY 334 AspProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyrAlaIleGlu 353  
 DB 1219 AATCCATTATGTCAGGAACCAAGCATCTCTTTGAAAGGCGATCTCTATATGCTATTGAA 1278  
 QY 354 ArgValLeuLysLeuSerValProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhe 373  
 DB 1279 AGAGTGTGAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCAATTTCTACTGCTTC 1338  
 QY 374 PheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyr 393  
 DB 1339 TTCCACCTTTGGTTAAACATATTTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATCTAC 1398  
 QY 394 LysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetProVal 413  
 DB 1399 AAGATTGTGGGAATGCAAAAGTGTGGGAGATTACTGGAGAATGTGGAATATGCTGTT 1458  
 QY 414 HisLysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIleProLysThr 433

XX

In the present invention, cDNA libraries from Arabidopsis, corn, rice, soybean and wheat were screened for sequences with homology to a putative acyl CoA cholesterol acyltransferase related gene from Arabidopsis thaliana and diacylglycerol acyltransferases from Homo sapiens and Mus musculus. The cDNA clones identified from this process were used to form complete diacylglycerol acyltransferase cDNA sequences. The present sequence is Arabidopsis thaliana diacylglycerol acyltransferase cDNA. Diacylglycerol acyltransferases are involved in the synthesis of triacylglycerols. Alteration of the expression of the diacylglycerol acyltransferase DNA can be useful for increasing the level of oils in plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful as herbicides.

XX Sequence 1888 BP; 457 A; 394 C; 429 G; 608 T; 0 other;

## Alignment Scores:

Pred. No.: 2,19e-252 Length: 1888  
Score: 2753.00 Matches: 518  
Percent Similarity: 99.62% Conservative: 0  
Best Local Similarity: 99.62% Mismatches: 2  
Query Match: 99.35% Indels: 0  
DB: 21 Gaps: 0

US-09-623-514A-2 (1-520) x AAA48932 (1-1888)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrValThrGluAsnGlyGlyGluPhe 20  
DB 131 ATGGCGATTTTGGATTCCTGCTGGCGTTACTACGGTGCAGGAGAACGGTGGCGGAGATTC 190  
QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40  
DB 191 GTCGATCTTGTAGAGCTGCTGCAGCGAAATCGATCGATCTTCTTAACGGACTTCCT 250  
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
DB 251 CTCCTGCTGGTTCGATTAATTAATCTCTCGGATGATGTTGGAGCTCCCGCCGACGTAGG 310  
QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
DB 311 GATCGGATGATTCGTTGTTAAACGATGACGCTCAGGGAACACCAATTTGGCCGAGAT 370  
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyAsn 100  
DB 371 AATAACGGTGGTGGCGTAATAACGGTGGTGGGAAGCGCGCGGAGGAAGGAAGGAAC 430  
QY 101 AlaAspAlaThrPheThrTyrrArgProSerValProAlaHisArgArgAlaArgGluSer 120  
DB 431 GCCGATGTACGTTTACGTATCGACCGTGGTTCAGCTCATCGGAGGCGGAGAGAGT 490  
QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
DB 491 CCACATTAGCTCCGACGCAATCTTCAACAGACGCCATGCCGGATTATTCACCTCTGTGTA 550  
QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrrGlyTrp 160  
DB 551 GTAGTCTTATTGCTGCTAAACAGTAGACTCATCATCGAAATCTTATGAAGTATGTTGG 610  
QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180  
DB 611 TTGATCAGAACGGATTTCTGGTTTATGTTCAAGATCGTGGAGATTTGGCCGCTTTTCATG 670  
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200  
DB 671 TGTGGATATCCCTTTCGATCTTCTTGGCTGCCCTTACGGTTGAGAAATGGTACTT 730  
QY 201 GlnLysTyrrIleSerGluProValIlePheLeuHisIleIleIleThrMetThrGlu 220  
DB 731 CAGAAATACATATCAGAACCTGTTGGCATCTTCTTCATATTTATCATCACCATGACAG 790  
QY 221 ValLeuTyrrProValTyrrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
DB 791 GTTTGTATCCAGTTTACCTACCTAACGGTGTGATCTGCTTTTATTCAGGTGTCACT 850

QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrrAlaHisThrSerTyrr 260  
DB 851 TTGATGCTCTCCTCACITTCGATTTGGTGGCTTAAAGTTGGTTCTTATGCTCATACTACTAT 910  
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrrVal 280  
DB 911 GACATAAGATCCCTAGCCAAATGAGCTGATAAGGCCAATCCTGAAGTCTCTACTACGTT 970  
QY 281 SerLeuLysSerLeuAlaTyrrPheMetValAlaProThrLeuCysTyrrGlnProSerTyrr 300  
DB 971 ACCTTGAAGAGCTTGGCATATTTTCATGGTGGCTCCCAATTTGGTTATCAGCAAGTTAT 1030  
QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle 320  
DB 1031 CCACGTTCTGCTATATACGAAGGCTTGGTGGCTCGTCAATTTTCAAAACTGGTCATA 1090  
QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrrIleAsnProIleValArgAsnSer 340  
DB 1091 TTCACCGGATTCATGGGATTTATAATAGAACAAATATATAATCTATTGTCAGGAATCA 1150  
QY 341 LysHisProLeuLysGlyAspLeuLeuTyrrAlaIleGluArgValLeuLysLeuSerVal 360  
DB 1151 AAGCATCTCTTGAAGGCGCATCTCTATATGCTATTGAAAGAGTGTGAAGCTTTCAGTT 1210  
QY 361 ProAsnLeuTyrrValTrpLeuCysMetPheTyrrCysPhePheHisLeuTrpLeuAsnIle 380  
DB 1211 CCAAAATTTATATGTTGGCTCTGCATGTTCTACTGCTTCTCCACCTTTGGTTAAACATA 1270  
QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrrLysAspTrpTrpAsnAlaLys 400  
DB 1271 TTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATCTACAAAGATTTGGTGAATGCANAA 1330  
QY 401 SerValGlyAspTyrrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420  
DB 1331 AGTGTGGGAGATTTACTGGAGATGTGGAATATGCCCTGTTTCATAAATGGATGGTTCGACAT 1390  
QY 421 IleTyrrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe 440  
DB 1391 ATATACTTCCCGTCTGTCGACGAGAACATACCAAGACACTGCCATTATCATTTGCTTTC 1450  
QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460  
DB 1451 CTAGTCTCTGCACTCTTTCATGAGCTATGATCGCAGTTCTTGTGCTCTCTCAAGCTA 1510  
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrrLeuGln 480  
DB 1511 TGGGCTTTTCTTGGGATTAATGTTTTCAGGTGCCCTTTGGTCTTCTCATCAAACTATCTACAG 1570  
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500  
DB 1571 GAAAGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTCATCTTCTGCTTTCGGA 1630  
QY 501 GlnProMetCysValLeuLeuTyrrTyrrHisAspLeuMetAsnArgLysGlySerMetSer 520  
DB 1631 CAACCGATGTGTGCTTCTTTTATTACCACGACCTGATGAACCGAAAGGATCGATGTCA 1690  
RESULT 6  
AAA51484  
ID AAA51484 standard; cDNA; 1985 BP.  
XX  
AC AAA51484;  
XX 09-OCT-2000 (first entry)  
DT  
XX  
DE A. thaliana AS1 diacylglycerol acyltransferase cDNA.  
XX  
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;  
KW size; weight; carbon flux; TAG1; insertion mutant; ss.  
XX  
OS Arabidopsis thaliana.  
XX  
FH Key Location/Qualifiers

QY 1 MetAlaIleLeuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20  
DB 237 ATGGCGATTGCGATTCTCTGCGCTTACTACGGTGACGAGAACGGTGGCGAGAGTTC 296  
QY 21 ValAspLeuAspArgLeuArgArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40  
DB 297 GTGCGATCTTGATAGGCTTCGTGACGGAATCGAGATCGGATCTTCTTAACGGACATCTT 356  
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
DB 357 CTCTCTGGTTCGATATAATCTCTCGGATGATGTTGGAGCTCCCGCGAGCTTAGG 416  
QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
DB 417 GATCGGATGATGATTCCTGTTAACGATGACGCTCAGGGAACAGCCAAATTTGGCCGGAGAT 476  
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGly 100  
DB 477 AATAACGGTGGTGGCATATAATACGGTGGTGGNAGAGCGCGGAGAGGAAGAAC 536  
QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120  
DB 537 GCGGATGCTACGCTTACGATACGACGCTCGGTTCCAGCTCATCGGAGGCGAGAGAGT 596  
QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
DB 597 CCACCTAGCTCCGACCAATCTCAACAGAGCCATGCCGGATATTCAACCTCTCTGTA 656  
QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyr 160  
DB 657 GTAGTCTTATTGCTGTAACAGTAGACTCATCATCGAATCTTATGAAGTATGTTGG 716  
QY 161 LeuIleArgThrAspPheThrPheSerArgSerArgSerLeuArgAspTrpProLeuPheMet 180  
DB 717 TTGATCAGAACGGATTCTTGGTTAGTTCAAGATCGCTCGAGATTGGCGCTTTTCATG 776  
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200  
DB 777 TGTGTATATCCCTTCGATCTTCTTGGTGGCTTACGGTTGAGAAATTTGGTACT 836  
QY 201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220  
DB 837 CAGAAATACATATCAGAACCTGTTGTCATCTTCTTCATATATATATACCATCAGACAG 896  
QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
DB 897 GTTTTGTATCCAGTTTACGTCACCTTAAGGTGTGATCTGCTTTTATCAGGTGTCACT 956  
QY 241 LeuMetLeuLeuThrCysIleValTyrLeuLysLeuValSerTyrAlaHisThrSerTyr 260  
DB 957 TTGATGCTCCTCCTCCTGATTTGGTGGCTAAAGTTGGTTTCTTATGCTCATACTAGTAT 1016  
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyrVal 280  
DB 1017 GACATAAGATCCCTAGCAATCGAGCTGATAGGCCAATCTTGAAGTCTCTACTACTAGT 1076  
QY 281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300  
DB 1077 AGCTTGAAGAGCTTGGCATATTTTCATGTCGCTCCACATTTGTTATCAGCAAGTAT 1136  
QY 301 ProArgSerAlaCysIleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuValIle 320  
DB 1137 CCACGTTCTGCAATGATATACGAAGGGTGGGTGGCTGCGCAATTTGCAAACTGGTCAVA 1196  
QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340  
DB 1197 TTCACGGATTCATGGATTTATATAGAACATATATAAATCTTATGTCAGGAAGTCA 1256  
QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360  
DB 1257 AAGCATCCCTTGAAGGCGATCTTCTATATGCTATTGAAAGAGTGTGAAGCTTTTCAGTT 1316

QY 361 ProAsnLeuTyrValTrpTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380  
DB 1317 CCAAAATTTATATGTGTGGCTGTCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATA 1376  
QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLys 400  
DB 1377 TTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATTTCTCAAGATTTGGTGAATGCAAAA 1436  
QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420  
DB 1437 AGTGTGGAGATTACTTGGAGAAATGCGAATATGCCCTGTTTCAAAATGGATGGTTCGACAT 1496  
QY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440  
DB 1497 ATATACTTCCGCTGCTTGGCAGCAAGATACCAAGACACTCCCAATATCATTTGCTTTC 1556  
QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460  
DB 1557 CTAGTCTCTGCACTCTTTCATGAGCTATGATCGAGTCTCTTGTCTCTCTTCAAGCTA 1616  
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480  
DB 1617 TGGGCTTTTCTTGGGATTTATGTTTTCAGGTGCTTGTCTTTCATCAAACTATCTACAG 1676  
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500  
DB 1677 GAAAGTGTGGCTCAACGGTGGGGAACATGATCTTCTGTTTCATCTCTGCAATTTTCGGA 1736  
QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520  
DB 1737 CAACCGATGTGTGCTTCTTTATTACCAGCACTCATGACCGAAGGATCGATGTCAT 1796  
RESULT 5  
-AAA48932  
ID AAA48932 standard; cDNA; 1888 BP.  
XX AC AAA48932;  
XX DT 06-DEC-2000 (first entry)  
XX DE Arabidopsis diacylglycerol acyltransferase cDNA.  
XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;  
XX KW triacylglycerol; herbicide; EC2.3.1.20; ss.  
XX OS Arabidopsis thaliana.  
XX FH Key Location/Qualifiers  
XX CDS 131..1693  
XX FT /tag- a  
XX FT /product= Diacylglycerol\_acyltransferase  
XX PN W0200032756-A2.  
XX PD 08-JUN-2000.  
XX PF 01-DEC-1999; 99WO-US28354.  
XX PR 02-DEC-1998; 98US-0110602.  
XX PR 31-MAR-1999; 99US-0127111.  
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.  
XX PI Cahoon EB, Kinney AJ, Cahoon RE;  
XX DR WPI: 2000-412308/35.  
XX DR P-PSDB; AAY94512.  
XX PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for  
XX PT synthesis of triacylglycerols and increasing the level of oils in plant  
XX PT seeds  
XX PS Claim 4; Page 38; 62pp; English.

Db 717 TTGATCAGACGGATTCTGCTTTAGTTCAAGATCGCTGCAGATGGCCCTTTTCATG 776  
 Qy 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200  
 Db 777 TGGTGTATATCCCTTTCGATCTTCTCCCTTTGGCTGCCCTTACGGTTGAGAAATGTGACTT 836  
 Qy 201 GlnIysTrpIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220  
 Db 837 CAGAAATACATPATCAGAACCTGTGTCTCATCTTCTTATATATATATATCACCATGACAG 896  
 Qy 221 ValLeuTrpProValTrpValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
 Db 897 GTTTTGTATCCAGTTTACCTCACCTAAGGTGATCTGCTTTTATCAGGTTCACT 956  
 Qy 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTrpAlaHisThrSerTrp 260  
 Db 957 TTGATGCTCCTCACTTGCATTTGTGGCTAAAGTTGGTTTCTTATGCTCATCTAGCTAT 1016  
 Qy 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTrpVal 280  
 Db 1017 GACATAAGATCCCTAGGCAATGCAGCTGATAGGCCAATCCTGAAGTCTCCTACTAGCTT 1076  
 Qy 281 SerLeuLysSerLeuAlaTrpPheMetValAlaProThrLeuCysTrpGlnProSerTrp 300  
 Db 1077 AGCTTGAAGAGCTTGGCATATTTTCATGCTGCTCCACATTTGTTATCAGCCAAAGTTAT 1136  
 Qy 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaAlaArgGlnPheAlaLysLeuValIle 320  
 Db 1137 CCACGTTCTGCATGTATACGGAAGGTTGGTGGCTGCTCAATTTGCNAAGTGGTCATA 1196  
 Qy 321 PheThrGlyPheMetGlyPheIleIleGluGlnTrpIleAsnProIleValArgAsnSer 340  
 Db 1197 TTCACGGATTCATGGGATTTATATAGAACATATATATAATCTATTTGTCAGGAAGTCA 1256  
 Qy 341 LysHisProLeuLysGlyAspLeuLeuTrpAlaIleGluArgValLeuLysLeuSerVal 360  
 Db 1257 AAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTGAAAGCTTTTCAGTT 1316  
 Qy 361 ProAsnLeuTrpValTrpLeuCysMetPheTrpCysPhePheHisLeuTrpLeuAsnIle 380  
 Db 1317 CCAAAATTTATATGTTGGCTGCTGATGTTCTTACTGCTCTTCCACCTTTGGTTAAACATA 1376  
 Qy 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTrpLysAspTrpTrpAsnAlaLys 400  
 Db 1377 TTGTCAGAGCTTCTCTGCTTGGGGATCGTGAATTTACAAAGATTTGGTGAATGCAAAA 1436  
 Qy 401 SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420  
 Db 1437 AGTGTGGAGATTAAGTGGAGATGGAATGCTGTTCAATGATGATGGTTCGACAT 1496  
 Qy 421 IleTrpPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440  
 Db 1497 ATATATTCGCTGCTTGGCAGCAAGATACCAAGACACTCGCCATATCATTTGCTTC 1556  
 Qy 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460  
 Db 1557 CTAGTCTCTGCAGTCTTTCATGAGTATGATCGACGTTCTTCTGCTCTCTTCAAGCTA 1616  
 Qy 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTrpLeuGln 480  
 Db 1617 TGGGCTTTCTTGGGATATGTTTTCAGGTGCTTTGGTCTTCATCAACAAACTATCTACAG 1676  
 Qy 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500  
 Db 1677 GAAAGTTTGGCTCAAGGGTGGGGAACATGATCTTCTGGTTCATCTCTCTGCTTTTCGGA 1736  
 Qy 501 GlnProMetCysValLeuLeuTrpTrpHisAspLeuMetAsnArgLysGlySerMetSer 520  
 Db 1737 CAACCGATGTGTGCTCTTCTTATACACGACCTGATGAACCGAAAGGATGATGATCA 1796  
 RESULT 4  
 ID AAS01106 standard; cDNA; 1942 BP.

XX AAS01106;  
 AC 31-MAY-2001 (first entry)  
 DT Arabidopsis thaliana sterol acyltransferase ACAT EST sequence.  
 DE Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;  
 KW acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;  
 KW nutritional supplement; dairy product; food product; salad dressing;  
 KW expressed sequence tag; EST; ss.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX Key Location/Qualifiers  
 FT CDS 237..1799  
 FT /\*tag= a  
 FT /product= "ACAT"  
 XX WO200116308-A2.  
 XX 08-MAR-2001.  
 XX 30-AUG-2000; 2000MO-US23863.  
 XX 30-AUG-1999; 99US-0152493.  
 XX (MONS ) MONSANTO CO.  
 PI Lassner M, Van Eenennaam A;  
 XX WPI; 2001-169010/17.  
 DR P-PSDB; AAU00462.  
 XX New isolated nucleic acid encoding plant lecithin:cholesterol  
 acyltransferase-like or acyl (coenzyme A) CoA:cholesterol  
 acyltransferase-like polypeptides, for modifying the sterol content and  
 oil production of plants -  
 Claim 4; Page 105-106; 127pp; English.  
 The present sequence encodes for Arabidopsis thaliana  
 acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel  
 polynucleotides encoding the plant sterol acyltransferases LCAT  
 (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341)  
 and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open  
 reading frame, LRO1 gene sequence (AAS01342), and a rat ACAT (AAS01105)  
 cDNA sequence are also described. The polynucleotides encoding LCAT  
 or ACAT are used to produce LCAT or ACAT polypeptides. They can also  
 be used in a recombinant construct to transform a host cell (preferably  
 of a plant) or a plant. The recombinant construct is used to increase or  
 decrease the sterol content of the host cell or plant. It can be used to  
 alter oil production of the cell or plant, preferably by increasing it.  
 The oil of the plant or the plant itself is used as a food product, or  
 as nutritional or dietary supplements, or in pharmaceutical compositions  
 for lowering cholesterol. The oil can be used in foods e.g. margarine,  
 butter, cooking oil, and dressings e.g. salad dressings, mayonnaise,  
 cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips,  
 baked goods, pastries, cookies, snack bars, confections, chocolates, and  
 beverages. The alteration in sterol content and/or composition can also  
 provide a plant with tolerance to stress and insect damage.  
 XX Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 4,39e-254 Length: 1942  
 Score: 2771.00 Matches: 520  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 Db: 22 Gaps: 0  
 US-09-623-514A-2 (1-520) x AAS01106 (1-1942)



DR P-PSDB; AAY54143.  
 XX Novel polynucleotides used for modifying plant oil composition and for  
 PT developing products for treating e.g. cancer, diabetes, cardiopulmonary  
 PT disease or metabolic disorders  
 XX  
 XX Claim 4; Fig 1; 89pp; English.  
 PS  
 XX The present sequence encodes an acyl-CoA:cholesterol acyltransferase  
 CC (ACAT) related protein. The ACAT-like protein is active in the formation  
 CC of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and  
 CC sterol and/or diacylglycerol substrate. The DNA can be used for  
 CC modifying the lipid composition of plant cells. The ACAT-like protein  
 CC has diacylglycerol acyltransferase (DGAT) activity, and so the  
 CC synthesis of triglycerides can be suppressed or increased using the  
 CC DNA. The protein can be used to produce plant oils with a modified  
 CC triglyceride content. The products can also be used to identify  
 CC antagonists and agonists of DGAT activity. Such agonists and  
 CC antagonists are particularly useful in treating or ameliorating  
 CC diseases associated with DGAT activity, including diseases associated  
 CC with altered cellular diacylglycerol concentration or PKC activity,  
 CC including cancer, diabetes, cardiopulmonary diseases e.g. heart failure,  
 CC atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,  
 CC metabolic disorders, obesity, diseases associated with abnormal lipid  
 CC metabolism, and diseases associated with abnormal fat absorption,  
 CC lipoprotein secretion and adipogenesis:  
 XX  
 SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;  
  
 Alignment Scores:  
 Pred. No.: 4,39e-254 Length: 1942  
 Score: 2771.00 Matches: 520  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0  
  
 US-09-623-514A-2 (1-520) x AAZ45371 (1-1942)  
  
 QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20  
 Db 237 ATGGCGATTTTGGATCTGCTGGCGTTACTACGGTGACGGAGACGGTGGCGGAGAGTTC 236  
  
 QY 21 ValAspLeuAspArgLeuArgArgLyssSerArgSerAspSerSerAsnGlyLeuLeu 40  
 Db 297 GTCGATCTTGATAGGCTTCGTGACGGGAAATCGAGATCGGATTCCTTAACGGGACTTCTT 356  
  
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
 Db 357 CTCCTCTGGTTCGGATAAATAATTCCTTCGGATGATGTGGAGTCCCGCCGACGTTAGG 416  
  
 QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
 Db 417 GATCGGATTGATTCGGTTGTAAACGATGACGCTCAGGGAACAGCCCAATTTGGCCGGAGAT 476  
  
 QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyAsn 100  
 Db 477 AATAACGGTGGTGGCGATAATAAACGGTGGTGGAGAGCGCGGAGAGGAGGAGGAAAC 536  
  
 QY 101 AlaAspAlaThrPheThrTyrrArgProSerValProAlaHisArgArgAlaArgGluSer 120  
 Db 537 GCGGATGCTACGTTTACGCTATCGACCGTCTCCAGCTCATCGAGGCGGAGAGAGT 596  
  
 QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
 Db 597 CCACCTAGCTCCGACGCAATCTTCAACAGAGCCATCCCGGATTTATTCAACCCCTGTGTA 656  
  
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysrGlyTrp 160  
 Db 657 GTAGTCTTATTGCTGTAACAGTAGACTCATCATCGAAATCTTATGAAGTAGTAGTGTGG 716  
  
 QY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180  
 Db

QY 501 GlnPrometCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520  
 DB 1639 CAACCGATGTGTGCTTCTTTATACACGACCTGATGAACCGAAAAAGGATCGATGTCA 1698

RESULT 2  
 ID AAA88835 standard; DNA; 1942 BP.  
 XX  
 AC AAA88835;  
 XX  
 DT 19-FEB-2001 (first entry)  
 XX  
 DE Arabidopsis acyl CoA:cholesterol acyltransferase DNA.  
 XX  
 KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;  
 KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;  
 KW hypolipemic; ss.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 237..1799  
 FT /\*tag= a  
 FT /EC\_number= "2.3.1.26"  
 XX  
 PW WO200061771-A2.  
 XX  
 PD 19-OCT-2000.  
 XX  
 PF 12-APR-2000; 2000WO-US09696.  
 XX  
 PR 12-APR-1999; 99US-0128995.  
 XX  
 PA (MONS ) MONSANTO CO.  
 XX  
 PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;  
 PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;  
 XX  
 DR WP1; 2000-665136/64.  
 DR P-PSDB; AAB19740.  
 XX  
 PT Genetically engineering the biosynthetic pathways in plants involved in  
 PT the accumulation of sterol compounds and tocopherol to produce  
 PT compounds for lowering the level of low density lipoprotein cholesterol  
 PT in blood serum -  
 XX  
 PS Disclosure; Page 57-58; 166pp; English.  
 XX  
 CC The present sequence is that of Arabidopsis thaliana DNA encoding  
 CC acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740). Sterol  
 CC O-acyltransferases such as ACAT catalyse the formation of  
 CC cholesterol esters from cholesterol and long chain fatty acids.  
 CC Recombinant constructs of the invention are used to alter the  
 CC biosynthesis and accumulation of sterols and tocopherols in  
 CC transgenic plants. Seeds of such plants may contain elevated  
 CC levels of sitosterol and/or its esters, and alpha-tocopherol, and  
 CC reduced levels of campesterol and campestanol and their esters.  
 CC The seeds may also contain the novel sterol brassicatanol. Oil  
 CC obtained from the seeds can be used in food and pharmaceutical  
 CC compositions to lower levels of low density lipoprotein cholesterol  
 CC in blood serum. ACAT enzymes can be used in the present invention  
 CC to produce elevated levels of phytosterol and/or phytostanol esters.  
 XX  
 SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Alignment Scores:  
 Pred. No.: 4,39e-254 Length: 1942  
 Score: 2771.00 Matches: 520  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 21 Gaps: 0

US-09-623-514A-2 (1-520) x AAA88835 (1-1942)

QY 1 MetalalleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20  
 DB 237 ATGGCGATTTGGATTCTGCTGGCGTTACTACGGTGACGGAGACGGTGGCGGAGATTC 296  
 QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40  
 DB 297 GTCGATCTTGATAGCTTCGTCGACGGAAATCGAGATCGGATTCCTCTAACGGACTTCTT 356  
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
 DB 357 CTCCTCGTTCCGATAAATTCCTCCGATGATGTTGGAGCTCCGCCGACCTTAGG 416  
 QY 61 AspArgileAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
 DB 417 GATCGGATTGATTCGGTTGTTAACGATCAGCTCAGGGAACAGCCAAATTTGCCGGAGAT 476  
 QY 81 AsnAsnGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 100  
 DB 477 AATAACGGTGGTGGCGATAAATACGGTGGTGAAGAGCGCGGAGAGAGAGAGAAAC 536  
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgArgAlaArgGluSer 120  
 DB 537 GCCGATGCTACGTTTACGATCGACCCGCTCGGTTCCAGCTCATCGGAGGGCGGAGAGAGT 596  
 QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
 DB 597 CCACCTAGCTCGGACCGCAATCTTCAACAGAGCCATGCCGATATATTCACCTCTGTGTA 656  
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleLeuGluAsnLeuMetLysTyrGlyTrp 160  
 DB 657 GTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAATCTTATCAAGTATGTTGG 716  
 QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180  
 DB 717 TTGATCAGAAGCGATTTCTGTTTAGTTCAAGATCGCTCGAGATTTGCCGCTTTTCATG 776  
 QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200  
 DB 777 TGTGTATATCCCTTTCGATCTTCTTGGTGCCTTTACGGTTGAGAAATTTGTACTT 836  
 QY 201 GlnLysTyrIleSerGluProValIlePheLeuHisIleIleIleIleThrMetThrGlu 220  
 DB 837 CAGAAATACATATCAGAACCTGTTGTCATCTTCTTCATATATATTATCACCATGACAG 896  
 QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
 DB 897 GTTTTGTATCCAGTTTACGTCACCCCTAAGGTGTGATTTCTGCTTTTATCAGGTGCTACT 956  
 QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrAlaHisThrSerTyr 260  
 DB 957 TTGATGTCCTCCTTGCATTTGTTGGCTAAAGTTGGTTTCTTATGCTCATACTAGCTAT 1016  
 QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTrpVal 280  
 DB 1017 GACATAAGATCCCTAGCCAATGCAGCTGATAAGGCCAATCCTGAAGTCTCTACTACGTT 1076  
 QY 281 SerLeuLysSerLeuAlaTyrPheMetValIalProThrLeuCysTyrGlnProSerTyr 300  
 DB 1077 AGCTTGAAGAGCTTGGCATATTTTCATGGTCCGCTCCCAATTTGTTGTTATCAGCAAGTTAT 1136  
 QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle 320  
 DB 1137 CCAGTCTGCTATGATACGGAAGGGTGGTGGCTGCTCAATTTGCAAACTGGTCTATA 1196  
 QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340  
 DB 1197 TTCACCGGATTCATGGGATTTAATAAGAACATATATAAATCTTATGTCAGGAACCTCA 1256  
 QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360  
 DB 1257 AAGCATCCTTTTGAAGGCGATCTTCTATATGCTATTTGAAGAGAGTGTGTAAGCTTTTCAGTT 1316

/\*tag- a  
/product= Diacylglycerol\_acyltransferase

W0200036114-A1.

22-JUN-2000.

16-DEC-1999; 99WO-CA01202.

17-DEC-1998; 98US-0112812.

(CANADA ) NAT RES COUNCIL CANADA.

Zou J, Taylor DC, Wei Y, Jako CC;

WPI; 2000-431592/37.

P-PSDB; AAY96853.

New DNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana for transforming plants and regulating seed oil content, fatty acid synthesis and seed oil acyl composition in commercial and crop plants

Claim 1; Page 64-65; 91pp; English.

This cDNA encodes Arabidopsis thaliana diacylglycerol acyltransferase (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating seed oil content, the ratio of diacylglycerol/triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil acyl composition, seed size/weight and carbon flux into other seed components in commercial and crop plants. The natural formation of triacylglycerols can be modified to increase the yield in commercial plant oils or modify their composition to achieve specific commercial improvements of plants and plant products.

Sequence 1904 BP; 458 A; 399 C; 431 G; 616 T; 0 other;

#### Alignment Scores:

Pred. No.: 4,26e-254 Length: 1904  
Score: 2771.00 Matches: 520  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 21 Gaps: 0

US-09-623-514A-2 (1-520) x AA51482 (1-1904)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20  
DB 139 ATGGCGATTGGATTCTGCTGGCGTTACTACGGTCACGGAGAACCGTGGCGGAGAGTTC 198  
QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40  
DB 199 GTCGATCTTGATAGGCTTCGTCACGCAATCGAGATCGGATCTCTTAACGGACTTC 258  
QY 41 LeuSerGlySerAspAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
DB 259 CTCCTGGTTCGGATAAATCTCTCGGATGATGTTGGAGCTCCCGCCGACGTTAG 318  
QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
DB 319 GATCGGATTGATTCGTTGTTAACGATGACGTCAGGGAACAGCCAAATTTGGCCGAGAT 378  
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyArgGlyAsn 100  
DB 379 AATAAGGTTGGTGGCGATAATAACGGTGGTGGGAAGCGCGGAGAGGAGGAAAC 438  
QY 101 AlaAspAlaThrPheThrThrArgProSerValProAlaHisArgAlaArgGluSer 120  
DB 439 GCGGATGCTACGTTTACGATACCGCTCGGTTCCAGCTCATCGGAGGCGGAGAGAGT 498  
QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
DB 499 CCACCTAGTCCGACCAATCTCAACAGAGCCATGCCGGATTATTCACACCTCTGTGTA 558

QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTrpGlyTrp 160  
DB 559 GTAGTTCTTATTGCTGTAAACAGTAGACTCATCATGAAATCTTATGAAGTATGTTGG 618  
QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180  
DB 619 TTGATCAGAACGGATTCTGGTTTAGTTCAAGATCGCTGCGAGATTGGCGCTTTTCATG 678  
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200  
DB 679 TGTGTATATACCTTTTCGATCTTTCTTGGCTGGCTTTACGGTTGAGAAATTTGGTACTT 738  
QY 201 GlnLysTrpIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220  
DB 739 CAGAAATACATATCAGAACCTGTTGTCTCTTCTCATATATTATTCACCATGACAGAG 798  
QY 221 ValLeuTrpProValTrpValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
DB 799 GTTTCTTATCCAGTTTACGTCACCTAAGGTGTGATTTCTGCTTTTATCAGGTGCAC 858  
QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTrpAlaHisThrSerTrp 260  
DB 859 TTGATGCTCTCTCACTTGCAATGGTGGCTAAAGTTGCTTCTTATGCTCATAGCTAT 918  
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTrpTrpVal 280  
DB 919 GACATAAGATCCCTAGCCAAATGAGCTGATAGGCAATCTCTCAAGTCTCTACTACGTT 978  
QY 281 SerLeuLysSerLeuAlaTrpPheMetValAlaProThrLeuLysCysTrpGlnProSerTrp 300  
DB 979 AGCTTGAAGAGCTTGGCATAATTTTCATGGTGGCTCCACATTTGTGTTATCAGCAAGTAT 1038  
QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaAlaArgGlnPheAlaLysLeuValIle 320  
DB 1039 CCACGTTCTGCATGATATACGGAAGGTTGGTGGCTGCTCAATTTGCAAACTGGTCATA 1098  
QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTrpIleAsnProIleValArgAsnSer 340  
DB 1099 TTCACCGGATTCATGGGATTTATAATAGAACAATATAATAATCTTATGTCAGGAAGTCA 1158  
QY 341 LysHisProLeuLysGlyAspLeuLeuTrpAlaIleGluArgValLeuLysLeuSerVal 360  
DB 1159 AAGCATCCCTTTGAAGGCGCATCTCTATATGCTATTTCAAGAGTGTGTAAGCTTTTCAGT 1218  
QY 361 ProAsnLeuTrpValTrpLeuCysMetPheTrpCysPhePheHisLeuTrpLeuAsnIle 380  
DB 1219 CCAAAATTTATATGTTGGCTGTCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATA 1278  
QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTrpLysAspTrpTrpAsnAlaLys 400  
DB 1279 TTGGCAGAGCTTCTCTGCTTGGGGATCGTGAATTTCTACAAGATTGGTGAATGCAAAA 1338  
QY 401 SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420  
DB 1339 AGTGTGGGAGATTACTGGAGATGTTGGAATATGCCCTGTTCAATAATGATGGTTCCACAT 1398  
QY 421 IleTrpPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleLeuAlaPhe 440  
DB 1399 ATATATCTCCCGTGGCTTGGCAGCAAGATACCAACACACTCCCATTTATCATGTTCTTC 1458  
QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460  
DB 1459 CTAGTCTCTGAGCTTTTCATGAGCTATGATCGCAGTTCCTTGTCTCTCTTCAAGCTA 1518  
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTrpLeuGln 480  
DB 1519 TGGGCTTTTCTTGGGATTTATGTTTTCAGGTGCCCTTTGCTTCTTCACTCAAACTATCTACAG 1578  
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500  
DB 1579 GAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTCACTCTGCTGCTTTCGGGA 1638

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 31, 2003, 04:59:59 ; Search time 377 Seconds  
(without alignments)  
3723.363 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTWGGGEF.....QPMCVLLYYHDLNMRKSGMS 520

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool/US09623514/runat\_29082003.152159.46/app\_query.fasta\_1.711  
-DB=N\_Geneseq\_19Jun03 -QPMF=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=DIOSUM62 -TRANS=human40.cd1  
-LIST=45 -LOCALIGN=200 -THR\_SCORE=ptc -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09623514 -CGN\_1\_1\_471 @runat\_29082003.152159.46 -NCPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N\_Geneseq\_19Jun03.\*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	100.0	2771	1904	21	AAA51482	A. thaliana diacyl
2	100.0	2771	1942	21	AAA88835	Arabidopsis acyl C
3	100.0	2771	1942	21	AZ45371	Acyl-CoA:cholester
4	100.0	2771	1942	22	AAS01106	Arabidopsis thalia
5	99.4	2753	1888	21	AAA48932	Arabidopsis diacyl
6	99.2	2747.5	1985	21	AAA51484	A. thaliana AS11 d
7	80.3	2225	5193	21	AAA51483	A. thaliana diacyl
8	80.0	2216.5	5339	21	AAA51485	A. thaliana AS11 d
9	65.7	1821	1942	21	AAA48939	Soybean diacylglyc
10	57.7	1597.5	1587	21	AAA48938	Rice diacylglycer
11	57.6	1596.5	1975	21	AAA48942	Wheat diacylglyce
12	46.2	1279	1559	21	AAA48935	Corn diacylglyce
13	45.5	1260	1281	21	AAA48933	Corn diacylglyce
14	37.1	1028.5	901	21	AAA48936	Corn diacylglyce
15	28.7	794	1976	21	AAA76169	Human ACAT Related
16	27.9	774	978	21	AAA48934	Corn diacylglyce
17	27.9	773.5	1650	21	AZ49452	Mouse diacylglyce
18	27.9	773.5	1766	21	AAA88846	Rat acyl CoA:chole
19	27.9	773.5	1766	21	AZ45385	Acyl-CoA:cholester
20	27.9	773.5	1766	22	AAS01105	Rat sterol acyltra
21	27.6	764.5	1732	24	AAD40352	Bovine DGAT1 CDNA
22	26.7	738.5	1521	19	AAV01533	Human acylcoenzyme
23	25.6	709.5	1895	21	AAA88842	Human acyl CoA:cho
24	25.2	698	1895	21	AZ45383	DNA encoding a pro
25	25.2	698	380	24	ABQ85322	Arabidopsis thalia
26	23.5	652.5	629	21	AAA51486	EST with homology
27	23.5	652.5	629	21	AZ49453	A. thaliana Diacyl
28	21.0	582	7490	23	ABL28620	Drosophila melanog
29	20.7	573.5	993	22	ABA08413	Human CDNA Seq ID
30	20.7	573.5	993	24	ABV83750	Human polyomucleot
31	19.3	535	1122	20	AAX30335	DNA encoding a hum
32	19.0	526.5	983	19	AAV01539	Human acylcoenzyme
33	19.0	526.5	470	24	ABL93444	Arabidopsis thalia
34	16.6	459	3996	23	ABL28621	Drosophila melanog
35	15.9	440	275	25	ABX33635	Human GDP-mannose
36	15.2	421	774	22	AAS27244	CDNA encoding nove
37	14.9	413.5	3649	19	AAV01536	Human acylcoenzyme
38	14.9	412.5	4011	15	AAQ63212	Acetyl coenzyme A:
39	14.7	407.5	452	24	ABL81686	Human ovarian canc
40	14.5	401.5	3650	19	AAT96368	Human acyl-coenzym
41	13.9	384	11771	24	AAD40350	Bovine DGAT1 gene.
42	13.7	378.5	2421	19	AAV01538	Acylcoenzyme A:cho
43	13.6	377	2421	19	AAT96370	Yeast acyl-coenzym
44	13.6	376	542	22	AAS27552	CDNA encoding nove
45	13.5	375	1607	21	AZ57359	Mouse acyl CoA:cho

ALIGNMENTS

RESULT 1  
AAA51482

ID AAA51482 standard; cDNA; 1904 BP.

AC AAA51482;

XX 09-OCT-2000 (first entry)

XX A. thaliana diacylglycerol acyltransferase cDNA.

DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;  
KW size; weight; carbon flux; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 139..1701

```
Db      624 GTTAGTCTATCACCACCGGTTCCCTGGTGTGCTCTGGCATCATACTCCATCATCTTC 683
QY      250 LeuLysLeuValSerTyrAlaHisThrSertyr-----AspIleArgSer 264
Db      684 CTCAGCTTTTCTCTACCGGATGTCAATCTGTGTGGTCCGCCAGGAGGTCAGAGGC 743
QY      265 LeuAlaAsnAlaAlaAspIys-----AlaAsnProGluValSerTyr 278
Db      744 AAGCTGTGTCTGCGAGGAAGAGTCAAGTGGGGTGTGCTGCCAGACACACTGTAAGCTAT 803
QY      279 -----TyrValSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyr 296
Db      804 CCGGACACCTGACCTACCGAGATCTATTACTTCTATCTTGTCTCTACTTTGTGTAT 863
QY      297 GlnProSerTyrProArgSerAlaCysIleArgLysGlyTyrPValAlaArgGlnPheAla 316
Db      864 GAACCTCAACTTTCCTCGATCCCGCCGAAATACGAAGCGCTTCTGCTACGGGGGTTCTT 923
QY      317 LysLeuValIlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIle 336
Db      924 GAGATGCTCTTTTCCACCGAGCTTCAAGTGGGGCTGATCCAGCAGTGGATGTCCTACT 983
QY      337 ValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyrAla-----IleGluArg 354
Db      984 ATCCAGAACTCCATGAAGCCCTCAAG--GACATGGACTATTCAGCAATCATTTAGGCGT 1040
QY      355 ValLeuLysLeuSerValProAsnLeuTyrValTyrLeuCysMetPheTyrCysPhePhe 374
Db      1041 CTCCTAAAGCTGGCGGTCCCCAACCATCTGATATGGCTCATCTTCTTATTTGGCTTTC 1100
QY      375 HisLeuTyrLeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLys 394
Db      1101 CACTCATGCTCAATGCTGTGGCAGAGCTCCTGCAGCTTTGGAGACCGCGAGTCTACAGG 1160
QY      395 AspTyrTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTyrAsnMetProValHis 414
Db      1161 GACTGGTGAATGCTGAGTGTACCTACTTTTGGCAGAACTGGAAATATCCCCGTGCAC 1220
QY      415 LysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeu 434
Db      1221 AAGTGTGCATCAGACACTTCTACAAGCCTATGCTCAGACTGGCGCAGCAACAATGGATG 1280
QY      435 AlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAlaValPro 454
Db      1281 GCCAGGACTGGGGTCTTTTGGCGTCAGCCTCTTCCATGAGTACCTAGTGAGCATTCOC 1340
QY      455 CysArgLeuPheLysLeuTyrAlaPheLeuGlyIleMetPheGlnValProLeuValPhe 474
Db      1341 CTGAGGATGTCGGCTCTGGGCATTCACGCCATGATGGCTCAGTCCCTCCCTGCCCTGG 1400
QY      475 IleThrAsnTyrLeuGlnArgPheGlySerThrValGlyAsnMetIlePheTyrPhe 494
Db      1401 ATTGTGAACCGCTC-----TTCCAAGGAACATATGGCAATGCAGCTGTGTGG--- 1448
QY      495 IlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAsp 511
Db      1449 GTGACACTCATATTGGCAACCGGTGCTGTGCTCATGTATGTCCACGAC 1499
```

Search completed: August 31, 2003, 08:27:19  
Job time : 403 secs

QY 225 ValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeu 244  
Db 660 GCAGCTGTGGCCCTACTGTTGAGTCTATCACCCTCAGTGGTTCGGTGTCTCGTGGCA 719  
QY 245 Thr---CysIleValTrpLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg 263  
Db 720 TCATACTCCATCATGTTCTCAAGCTTTATTCCTACCGGATGTCACCTGTGGTGGCCG 779  
QY 264 -----SerLeuAlaAsnAlaAalaasp 270  
Db 780 CACGGAAGGTCGAAGCCAAAGCTGTCTCTACAGGGAAGAGTCAAGTGGGGTGTCT--- 836  
QY 271 LysAlaAsnProGluValSerTyr-----TyrValSerLeuLysSerLeuAlaTyrPhe 288  
Db 837 ---GCCAGCAAGCTGTGAGCTATCCAGACAACCTGACCTACCGAGATCTCTATTACTTC 893.  
QY 289 MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLys 308  
Db 894 ATCTTTGCTCTACTTTGTTGTTATGAACCTCAACTTCTCGTCCCGCCGCAATACGAAG 953  
QY 309 GlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle 328  
Db 954 CGCTTCTGTACGACGAGTCTTGAGATGCTCTTTTACCAGCTTCGAAGTGGGGCTG 1013  
QY 329 IleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu 348  
Db 1014 ATCCAACAGTGGATGTCCTACTATCCACAACCTCCATGAAGCCCTCAAG---GATATG 1070  
QY 349 LeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrp 366  
Db 1071 GACTATTACGGGATCATGTAGCGCTCTTAAAGCTGGCGGTCCCAACCATCTGTATGCTG 1130  
QY 367 LeuCysMetPheTyrCysPheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCys 386  
Db 1131 CTATCTTCTTATGTTTTCACCTGCTGCTCAATGCTGTGGCAGAGTCTCTGCAG 1190  
QY 387 PheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrp 406  
Db 1191 TTTGGAGACCGCGAGTCTACAGAGATGTTGTAATGCTGAGTCTGTCACTACTTTGG 1250  
QY 407 ArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCysLeu 426  
Db 1251 CAGAATGGGAATATCCCGTGCACAAAGTGTGTGCACTCAGACACTCTACAAGCCTATGCTC 1310  
QY 427 ArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuValSerAlaValPhe 446  
Db 1311 AGACATGGCAGCAGCAATGGTGGCCAGCAGGAGTATTTTGACCTCAGCTTCTTC 1370  
QY 447 HisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIle 466  
Db 1371 CATGAGTACCTAGTAGAGCGTTCCTCGCGATGTTCCGCTCTGGGCATTTCACAGCCATG 1430  
QY 467 MetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThr 486  
Db 1431 ATGGCTCAGTCCCTCAGTGGCTGGATGTGGGCGCATTC-----TTCCAAGGGAAC 1481  
QY 487 ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeu 506  
Db 1482 TATGGCAATGCAGCTGTGG---GTGACACTCATCATTTGGGCAACGGTGGCTGTGCTC 1538  
QY 507 LeuTyrTyrHisAsp 511  
Db 1539 ATGATGTCCACGAC 1553

## RESULT 15

US-10-157-855-15  
; Sequence 15, Application US/10157855  
; Publication No. US20020170091A1  
; GENERAL INFORMATION:  
; APPLICANT: Lassner, Michael W.  
; APPLICANT: Ruzinsky, Diane M.  
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic

; TITLE OF INVENTION: Acid Sequences  
; FILE REFERENCE: 16516.158  
; CURRENT APPLICATION NUMBER: US/10/157,855  
; CURRENT FILING DATE: 2002-05-31  
; PRIOR APPLICATION NUMBER: 09/326,203  
; PRIOR FILING DATE: 1999-06-04  
; PRIOR APPLICATION NUMBER: 60/088,143  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/108,389  
; PRIOR FILING DATE: 1998-11-12  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 15  
; LENGTH: 1766  
; TYPE: DNA  
; ORGANISM: Rattus sp.  
; US-10-157-855-15

## Alignment Scores:

Pred. No.: 2,7e-76 Length: 1766  
Score: 773.50 Matches: 193  
Percent Similarity: 53.58% Conservative: 84  
Best Local Similarity: 37.33% Mismatches: 171  
Query Match: 27.91% Indels: 69  
DB: 13 Gaps: 17

US-09-623-514A-2 (1-520) x US-10-157-855-15 (1-1766)

QY 14 GluAsnGlyGlyGluPheValAspLeuAspArgLeuArgArgLysSerArgSer 33  
Db 99 CAGGGTGGTAGTGGGCCCATGTAGACGAAGAGAGGTGCAGACGCCGTGTGGGCC 158  
QY 34 AspSerSerAsnGlyLeuLeuSerGlySerAspAsnAsnSerProSerAspVal 53  
Db 159 GAC-----TTGGCGCGCGGGGTGACCTCCGGTCCGGT---CCGGT 200  
QY 54 GlyAlaProAlaAspValArgAspArgIleAspSerValValAsnAspAlaGlnGly 73  
Db 201 CCGGCTCCAGCCACACCCCGGACAAA-----GACCGCAGACC 239  
QY 74 ThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGlyAspAsnGlyGlyArgGly 93  
Db 240 AGCGTG-----GCGCAGCGC 254  
QY 94 GlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSerValProAla 113  
Db 255 CACTGGGAGCTGAGG-----TGC 272  
QY 114 HisArgAlaArgGluSerProLeuSerSerAspAlaIlePheLysGlnSerHisAla 133  
Db 273 CAT---CGTCTGAAGACTCTTTGTCAGCTCAGACGAGCGGTTTC---AGCAATTACCGT 326  
QY 134 GlyLeuPheAsnLeuCysValValLeuIleAlaValAsnSerArgLeuIleGlu 153  
Db 327 GGATCTGAATTTGGTGGTGGTGGTCTGATCTGAGTAAATGCAAGGTTATTTTAGAG 386  
QY 154 AsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerArgSerLeu 173  
Db 387 AATCTTATCAAGTATGGCATCTCTGGTG---GATCCCATCCAGGTGCTCTCTGTTCTG 443  
QY 174 ArgAsp-----TrpProLeuPheMetCysIleSerLeuSerIlePheProLeu 190  
Db 444 AAGGACCCCTACAGCTGGCCTGCCCCATGCTTGCATCATTCATCAATAATCTTTATGTG 503  
QY 191 AlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValIle 210  
Db 504 GCTACATTTTCAGATTGAGAAAGCGCTGTGAGTGGTGGTCCCTGACAGAGAGATGGGCTG 563  
QY 211 PheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArg 230  
Db 564 CTGCTACATGTTGTTAACTTACCTGGCCACAATATCTGCTTCCACAGCTGTGGCTTACTG 623  
QY 231 CysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThr---CysIleValTrp 249

Db 660 GCAGCTGGCGCTTACTGTTGAGTCTATCACTCCAGTGGGTCCGTGTTGCTCGGCA 719  
QY 245 Thr---CysileValtrpLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg 263  
Db 720 TCATACCTCATCATGTTCTCAAGCTTATCTACCGGGGATGCAACCTGTGGTCCCG 779  
QY 264 -----SerLeuAlaAsnAlaAlaAsp 270  
Db 780 CAGCGAAGGGTCAAGCCAAAGCTGTCTACAGGGAAGAGGTCAAGTGGGCTGCT--- 836  
QY 271 LysAlaAsnProGluValSerTyr-----TyrValSerLeuLysSerLeuAlaTyrPhe 288  
Db 837 ---GCCACGACCTGTGAGCTATCCAGACACCTGACCTACCGAGATCTCTATTAATTC 893  
QY 289 MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLys 308  
Db 894 ATCTTTGCTCCTACTTGTGTTATGAACCTCAACTTCTCCGTCGCCCGCAATACGAAG 953  
QY 309 GlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle 328  
Db 954 CCGTTTCTGCTACGACGAGTCTTTCAGATGCTCTTTTACCGAGCTTCAAGTGGGGCG 1013  
QY 329 IleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu 348  
Db 1014 ATCCACAGTGTGATGCTCTACTATCCACACTCCATGAGCCCTTCAAG---GATARG 1070  
QY 349 LeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrp 366  
Db 1071 GACTATTTCAGGATCATTCAGCGTCTCTTAAAGCTGGCGTCCCAACCATCTGATCTGG 1130  
QY 367 LeuCysMetPheTyrCysPheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysCys 386  
Db 1131 CTATCTCTCTTCTTATGTTGTTTTCACCTCTCTCTCAATGCTGTGTCAGAGCTTCTG 1190  
QY 387 PheGlyAspArgGluPheTyrLysAspTrpTyrAsnAlaLysSerValGlyAspTyrTrp 406  
Db 1191 TTTGGAGACCGGAGTCTACAGAGATTGTTGATGCTGTGATGCTGTGCTACCTACTTTGG 1250  
QY 407 ArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCysLeu 426  
Db 1251 CAGAACTGGAATATCCCGTCGACAAAGTGGTGCATCAGACACTTCTACAGGCTATGCTC 1310  
QY 427 ArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuValSerAlaValPhe 446  
Db 1311 AGACATGGCAGCAGCAATGGGTGGCCAGCAGGAGTATTTTACCTCAGCGCTTCTTC 1370  
QY 447 HisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIle 466  
Db 1371 CATGAGTACTAGTAGGCTTCCCTGCGGATGTTCCGCTCTGGGCATTCACAGCCATG 1430  
QY 467 MetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThr 486  
Db 1431 ATGGCTCAGGTCCCACTGGCTGGATTGTGGCGGATTC-----TTCCAAGGGAAC 1481  
QY 487 ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeu 506  
Db 1482 TAGGCAATGCAGCTGTGTGG---GTGACACTCATCATTTGGCAACCGGTGTGCTGTC 1538  
QY 507 LeuTyrTyrHisAsp 511  
Db 1539 ATGATGTCACGAC 1553

## RESULT 14

US-10-273-438-9  
; Sequence 9, Application US/10273438  
; Publication No. US20030072757A1  
; GENERAL INFORMATION:  
; APPLICANT: Farese, Robert V.  
; APPLICANT: Cases, Sylvaine  
; APPLICANT: Smith, Steven  
; APPLICANT: Erickson, Sandra  
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase  
; FILE REFERENCE: UCAL-105C1P2

; CURRENT APPLICATION NUMBER: US/10/273,438  
; CURRENT FILING DATE: 2002-10-16  
; PRIOR APPLICATION NUMBER: US/10/040,315  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 60/107,771  
; PRIOR FILING DATE: 1998-11-09  
; PRIOR APPLICATION NUMBER: PCT/US98/17883  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: 09/103,754  
; PRIOR FILING DATE: 1998-06-24  
; PRIOR APPLICATION NUMBER: 09/339,472  
; PRIOR FILING DATE: 1999-06-23  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 9  
; LENGTH: 1650  
; TYPE: DNA  
; ORGANISM: mus musculus  
; US-10-273-438-9  
Alignment Scores:  
Pred. No.: 2,42e-76 Length: 1650  
Score: 773.50 Matches: 191  
Percent Similarity: 51.81% Conservative: 81  
Best Local Similarity: 36.38% Mismatches: 178  
Query Match: 27.91% Indels: 75  
DB: 14 Gaps: 15  
US-09-623-514A-2 (1-520) x US-10-273-438-9 (1-1650)  
QY 8 GlyValThrThrValThrGluAsnGlyGlyGlyGluPheValAspLeuAspArgLeuArg 27  
Db 141 GGCTCGCGGTTTCCCTCCAGGGTGTAGTGGCCCAAGGTAGAACAGGACGAGGTGCGA 200  
QY 28 ArgArgLysSerArgSerAspSerSerAsnGlyLeuLeuLeuSerGlySerAspAsn 47  
Db 201 GACGGCGTGTGAGCCCCGAC-----TTGGCGCCGGGGTGACGGCGG 245  
QY 48 SerProSerAspValGlyAlaProAlaAspValArgAspArgIleAspSerValVal 67  
Db 246 GCTCCGCGCTCG-----GCTCCAGCCCATACCCGGGACAA----- 281  
QY 68 AsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyAspAsn 87  
Db 281 ----- 281  
QY 88 AsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 107  
Db 282 ---GACGGGGGACGAGCGTGGCGGCTACTGGGATCTGAGGTGC----- 326  
QY 108 ArgProSerValProAlaHisArgArgAlaArgGluSerProLeuSerSerAspAlaIle 127  
Db 327 -----CATCGCTCTCAAGATTCCTTTGTTCAAGTCTCAGACAGTGGT 365  
QY 128 PheLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValValLeuLeuAlaValAsn 147  
Db 366 TTC---AGCAATATATCGTGTATCTCTGAATTTGGTGTGTGTGTGTGTGTGTGTGTGT 422  
QY 148 SerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrpIleLeuLeuArgThrPheTrp 167  
Db 423 GCAAGGTTATTTTATAGAACCTTATCAAGTATGTCATCTCTGGT---GATCCTATCCAG 479  
QY 168 PheSerSerArgSerLeuArgAsp-----TrpProLeuPheMetCysCysIleSer 184  
Db 480 GTGGTGTCTCTGTTTGAAGGACCCCTACAGCTGGCCTGCCCTGCCCTGATGATATGCA 539  
QY 185 LeuSerIlePheProLeuAlaAlaPheThrValGluLeuValLeuGlnLysTyrIle 204  
Db 540 TCCAAATATTTTGT 599  
QY 205 SerGluProValIlePheLeuHisIleIleIleIleIleIleIleIleIleIleIleIle 224  
Db 600 ACAGACGATGGGCTGCTGCTACATGTGGTTAACTGGCCCAACATCATTTGCTTCCCA 659

```
Db 586 CTGGCGCTGATGGCGGCACACCATCTCTCTCAAGCTCTTCTCTCTACCGC----- 636
Qy 260 TyrAspIleArgSer-----LeuAlaAsnAlaAspLys 271
Db 637 ---GACGTCACACTCATGTGTGGCGGAGCCAGGCGCTGCTCTGCGAGGAGAG 693
Qy 272 AlaAsnPro-----GluValSerTyr-----TyrValSerLeuLysSer 284
Db 694 GCCAGCAGTGTGTGCGCCGCACACACCGTGTGAGTACCGGACAACTCTGACCTACCGCAT 753
Qy 285 LeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAla 304
Db 754 CTCCTACTCTCTCTCTGCGCCCGCCACCTGTGCTACGAGCTCAACTTTCCCGCTCTCCC 813
Qy 305 CysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPhe 324
Db 814 CGCATCCGGAAGCGCTTCTCTCGGACGGATCTTGAGATGCTGTCTTCAACCGAGTC 873
Qy 325 MetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeu 344
Db 874 CAGGTGGGCTGTATCCAGCAGTGTGTGCTCCACCACTCCAGAACTCCATGAAGCCCTTC 933
Qy 345 LysGlyAspLeuLeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsn 362
Db 934 AAG---GACATGGACTACTACGCATCATCGAGCGCTCTCTGAAGCTGGCGGTCCCAAT 990
Qy 363 LeuTyrValTyrLeuCysMetPheTyrCysPheHisLeuTyrLeuAsnIleLeuAla 382
Db 991 CACCTCATCTGGCTCATCTTCTCTACTGCTCTTCTCACTCTCTGCTGCTGAATGCCGTGCT 1050
Qy 383 GluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAlaLysSerVal 402
Db 1051 GAGCTCATGAGTGTGGAGACGGGAGTCTACCGGGACTGGTGAAGTCCGAGTCTGTC 1110
Qy 403 GlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyr 422
Db 1111 ACCTACTTCTGGCAGAACTGGAACATCCCTGTGCACAAAGTGTGTCATCAGACACTTCAC 1170
Qy 423 PheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuVal 442
Db 1171 AAGCCCATCTCGACGGGCGAGCAGCAAGTGGATGGCCAGCAGGCGTGTCTCGGCC 1230
Qy 443 SerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAla 462
Db 1231 TCGGCTTCTTCCACGAGTACTGTGTGAGCGTCTCTGCGAATGTTCCGCTCTGGGCT 1290
Qy 463 PheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTyrLeuGlnGlu 481
Db 1291 TTCAGGGCATGATGGCTCAGATCCCACTGGCCCTGGTTCGTGGCGCGCTTTTTCAGAGGC 1350
Qy 482 ArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGln 501
Db 1351 AACTAT-----GGCAACGACGCTGTGG---CIGTGCCTCATCATCGGACAG 1395
Qy 502 ProMetCysValLeuLeuTyrTyrHisAsp 511
Db 1396 CCAATAGCGTCTCATGTACGTCCACGAC 1425
```

## RESULT 13

```
US-10-273-438-3
; Sequence 3, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farsese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
```

```
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: mus musculus
US-10-273-438-3

Alignment Scores:
Pred. No.: 2,42e-76 Length: 1650
Score: 773.50 Matches: 191
Percent Similarity: 51.81% Conservative: 81
Best Local Similarity: 36.38% Mismatches: 178
Query Match: 27.91% Indels: 75
DB: 14 Gaps: 15

US-09-623-514A-2 (1-520) x US-10-273-438-3 (1-1650)
```

```
Qy 8 GlyValThrValThrGluAsnGlyGlyGluPheValAspLeuAspArgLeuArg 27
Db 141 GCCTCGCGGGTTCCTCCAGGGTGTAGTGGCCCAAGGTAGAGACAGACAGGTGGCA 200
Qy 28 ArgArgLysSerArgSerAspSerSerAsnGlyLeuLeuLeuSerGlySerAspAsnAsn 47
Db 201 GACGGCGGTGTGAGCCCGAC-----TTGGCGCGCGGGGTGACGCGCG 245
Qy 48 SerProSerAspValGlyAlaProAlaAspValArgAspArgIleAspSerValVal 67
Db 246 GCTCGCGGTCCG-----GCTCCAGCCCATACCGGACAAA----- 281
Qy 68 AsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGlyAspAsn 87
Db 281 ----- 281
Qy 88 AsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 107
Db 282 ---GACGGCGGACGACGCGTGGCGGCTACTGGATCTGAGGTG----- 326
Qy 108 ArgProSerValProAlaHisArgArgAlaArgGluSerProLeuSerSerAspAlaIle 127
Db 327 -----CATGCTCTGCAAGATTCTTTGCTCAGCTCAGACAGTGGT 365
Qy 128 PheLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValValIleAlaValAsn 147
Db 366 TTC---ACCAATATFCGTGGTATCTGAATGGTGTGGTGTGGTGTGGTGTGGTGTGGT 422
Qy 148 SerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrLeuIleArgThrAspPheTrp 167
Db 423 GCAAGGTATTTTAGAGAACCTTATCAAGTATGCAATCCTCTGGTGTGGTGTGGTGTGGT 479
Qy 168 PheSerSerArgSerLeuArgAsp-----TrpProLeuPheMetCysCysIleSer 184
Db 480 GTGGTGTCTCTGTTTGTGAAGGACCCCTACAGCTGGCTGGCTGGCTGGCTGGCTGGCT 539
Qy 185 LeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTrpIle 204
Db 540 TCCAATATTTTGTGGCTGCATTTTCAGATTGAGAAGCCCTGGCCTGGCCTGGCCTGGC 599
Qy 205 SerGluProValIlePheLeuHisIleIleIleIleIleIleIleIleIleIleIleIle 224
Db 600 ACAGACAGATGGGCTGCTGCTACATGTTACCTGGCCACACATATTGCTTCCCA 659
Qy 225 ValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeu 244
Db ----- 244
```





```
Db 72 GCTGCATATGGAATATATGTCGATCCGAGATATGAAGATCAACCTTTAAAGTCTA 131
Qy 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305
Db 132 GTGTACTTCATGTTGGCCCAACACTTGTACCAAGCACTTATCTCAACTACATGT 191
Qy 306 IleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325
Db 192 ATTAGAAGGGTTGGTGACCCAGCACTCATATAAGTCCGTTGGNNNNNACAGGCTTGATG 251
Qy 326 GlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345
Db 252 GGCTTCATATATGAGCAATATATAACCAATTTGTAAGATTCCTCAACATCCACTGAAA 311
Qy 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365
Db 312 GGGANNNNNGAATGCTATAGAAGAGTCTTAACATCTCAGTGCCCAACATATATATGTA 371
Qy 366 TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeu 385
Db 372 TGGCTTTGCATGTTCTATGTCNNNNNCATTTATGCTGAACATTTGAGCTGAACCTCCTC 431
Qy 406 CysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyr 405
Db 432 TGTTCGCTGACCGTGAATCTATAGGACTGGTGAATGCCAAACTGTTGAAGAGTAC 491
Qy 406 TrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCys 425
Db 492 TGGAGGATGTGAACATGCGCTGTTCAATAGTGATCATCAGACACATATATTTCCATGT 551
Qy 426 LeuArgSerLysIleProLysThrLeuAlaIleIleLeuAlaPheLeuValSerAlaVal 445
Db 552 ATAAGGAAAGGCTTTTCAGGGGTGAGTATTAATCTCGTTCTGTTTTCAGCTGTA 611
Qy 446 PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly 465
Db 612 TTCATGAGATATATATGCGGTGCCGTCACATTTTCAAAATCTGGGCATTTTCTGGG 671
Qy 466 IleMetPheGlnValProLeuValPheIlePheTrpAsnTyrLeuGlnArgPheGlySer 485
Db 672 ATCATGTTTCAGATACCGTGGTATTTCTTGACAGATATCTCCATGTCAGTCAAGCAT 731
Qy 486 Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCys 504
Db 732 GTAATGTTGGCAACATGATATTTTGTTC---TTCAGTATAGTCGACAGCCGATGTGT 788
Qy 505 ValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
Db 789 GTCCTTCTACTACTACCATGAGTGCATGAACAGGCGAGCCCGCAAGT 836
```

## RESULT 11

```
US-10-278-733-2
; Sequence 2, Application US/10278733
; Publication No. US20030100480A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Steven
; APPLICANT: Farese, Robert V Jr
; TITLE OF INVENTION: Methods and compositions for modulating
; TITLE OF INVENTION: sebaceous glands
; FILE REFERENCE: UCAL-105CIP4
; CURRENT APPLICATION NUMBER: US/10/278,733
; PRIOR FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
```

Alignment Scores:  
Pred. No.: 1,9e-78 Length: 1467  
Score: 791.50 Matches: 194  
Percent Similarity: 53.14% Conservative: 77  
Best Local Similarity: 38.04% Mismatches: 174  
Query Match: 28.56% Indels: 65  
DB: 14 Gaps: 17

US-09-623-514A-2 (1-520) x US-10-278-733-2 (1-1467)

Qy 25 ArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeuSerGlySer 44  
Db 22 CGGCGCGGAGACAGGTCGCGCCCTCGAGCCACCGCGGC----- 63  
Qy 45 AspAsnAsnSerProSerAspValIcylAlaProAlaaspValArgAspArgIleasp 64  
Db 64 -----GGCGGGCTCGCGCGCGGCGGAGAGGAG----- 90  
Qy 65 SerValValAsnAspAspAlaGlnGly---ThrAlaAsnLeuAlaGlyAspAsnGly 83  
Db 91 -----GTGCGGAGCGCGCTGCGGCGCCCGACGTCGGAGCGCGGGGAGCCGCCAGCC 144  
Qy 84 GlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 103  
Db 145 CCGGCCCCCAACAGGAGCGGAGCGCGCGTGGCGAGCGGCCACTGGAGCTCAGGTGC 204  
Qy 104 ThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSerProLeuSer 123  
Db 205 -----CATGCGCTGCAGGATCTTTATTTCAGC 231  
Qy 124 SerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValValLeu 143  
Db 232 TCTGACAGTGGCTTC---AGCAACTACCGTGGCATCTGAACTGGTGTGTGTGTGTGTG 288  
Qy 144 IleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrLeuIleArg 163  
Db 289 ATCTTGAGCAATGCGCGTTATTTCTGGAGAACCTCATCAAGTATGTCATCTCGTGTG 345  
Qy 164 ThrAspPheTrpPheSerArgSerLeuArgAsp-----TrpProLeuPheMet 180  
Db 346 GACCCCATCCAGGTGGTTCTTCTTCTTGAAGATCCCTATAGTGGCGGCCCATGTC 405  
Qy 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200  
Db 406 CTGCTTATTTGGGCCAATGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 465  
Qy 201 GlnLysTyrIleSerGluProValIlePheLeuHisIleIleIleIleIleIleIle 220  
Db 466 GTGGTGGCTGACGAGGAGCGGCGGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 525  
Qy 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
Db 526 CTGTGTTTCCAGCGGT 585  
Qy 241 LeuMetLeuLeuThr---CysIleValTrpLysLeuValSerTyrAlaHisThrSer 259  
Db 586 CTGGCGGTGATGGCGCACACCATCTCTCTTCAAGCTCTCTCTCTCTCTCTCTCTCT 636  
Qy 260 TyrAspIleArgSer-----LeuAlaAsnAlaAlaAspLys 271

Score:	1279.00	Matches:	232
Percent Similarity:	65.50%	Conservative:	49
Best Local Similarity:	54.08%	Mismatches:	69
Query Match:	46.16%	Indels:	79
DB:	14	Gaps:	3

US-09-623-514A-2 (1-520) x US-10-223-076-17 (1-1572)

Qy	166	PheTrrPheSerArgSerLeuArgAspTrrPProLeuPheMetCysCysIleSerLeu	185
Db	1	TTTTGGTTTAATGCTFACATCATTCGGAGACTGGCCACTGCTAATGTTGGCTTAGCTCTA	60
Qy	186	SerIlePheProLeuAlaAaphethrValGlulysLeuValLeuGlnIlySer	205
Db	61	CCCATATTTCCCTTGGTGCATTTGCGATCGAAGTTGGCATTCAACAATCTCGTTAGT	120
Qy	206	GlulProValValIlePheLeuHisIleIleIlethrMetThrGluValLeuTyrProVal	225
Db	121	GATCCTTGCTACCTGTTTTCACATCCTTTTACAACATTTGAAATTTGATATCATCAGTG	180
Qy	226	TyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThr	245
Db	181	CTCGTGATCTTAAGTGTGATCTGCAGTTTATCAGGCTTTGTGTGATGTTTATGGC	240
Qy	246	CysIleValTrrPLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArgSerLeu	265
Db	241	TGCATTTGTTGGCTGAAGCTTGATCTTTTGCACATACAAACCATGATATATAAGAAAC	300
Qy	266	AlaAsnAlaAlaAspLysAlaAsnProGluValSer-----Tyr	278
Db	301	ATCACAAAGCGCAAGAAGGTGTGAATGAAGTACCGCGCTGGCATAGATAATTTACAA	360
Qy	279	TyrValSerLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnPro	298
Db	361	GCTCCAACCTCTGGGAGCTTAACATCTTCATGATGGCTCGACACTCTGTATACAGCA	420
Qy	299	SerTyrProAsnSerAlaCysIleArgLysGlyTrrPValAlaAlaArgGlnPheAlaLysLeu	318
Db	421	AGTTATCTCTCGAACACCTTATGTAGAAAAGTTGGCTGGTGGCTCAAGTATTCTATAC	480
Qy	319	ValIlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArg	338
Db	481	TGTATATTACTGGTCTCCAAGGATTCATTATGAGCAATACATAATCCTATGTGTGG	540
Qy	339	AsnSerLysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeu	358
Db	541	AACTCTCAACATCCATTGATGGAGGATTACTGAATGCTGTAGAGACTGTTTGAAGCTC	600
Qy	359	SerValProAsnLeuTyrValTrrPLeuCysMetPheTyrCysPheHisLeuTrrPLeu	378
Db	601	TCATTAACCAATGTCTACCTGGCTTGGCATGTTTATTGCTTTTCCATCTGTGTTTA	660
Qy	379	AsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrrPrrAsn	398
Db	661	AACACTTGTGCGAATTTCTCGATTTGGTGACCGAGAATTCACAAAGACTGGTGAAT	720
Qy	399	AlaLysSerValGlyAspTyrTrrPArgMetTrrPAsnMetProValHisLysTrrPMetVal	418
Db	721	GCMAAGACAAATTGATGAGTACTGGGAATAATGGACATGCTGTGCATAAATGGATGTT	780
Qy	419	ArgHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIle	438
Db	781	CGTCATATATATTTTCCCTTGGCATCGAAATGGTATATCAAGGAAGTGTGCTGTTTATA	840
Qy	439	AlaPheLeuValSerAlaVal-Phe-----	446
Db	841	TCGTTCTTGTGTTCTGCTGCTACTTCATGAGTAACCTTATTACTTTTTTCACTCTCTCAT	900
Qy	446	-----	446
Db	901	GCATATATTAATATATAGTCTCTATTTTCAAAATGTGCTCTTCGAGTTTTCGACATGCT	960
Qy	446	-----	446

; APPLICANT: Farese, Robert V  
 ; APPLICANT: Cases, Sylvaine  
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
 ; FILE REFERENCE: UCAL-105CIP3  
 ; CURRENT APPLICATION NUMBER: US/10/223,076  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 10/040,315  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 09/339,472  
 ; PRIOR FILING DATE: 1999-06-23  
 ; PRIOR APPLICATION NUMBER: 60/107,771  
 ; PRIOR FILING DATE: 1998-11-09  
 ; PRIOR APPLICATION NUMBER: PCT/US98/17883  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: 09/103,754  
 ; PRIOR FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 6  
 ; LENGTH: 1446  
 ; TYPE: DNA  
 ; ORGANISM: Brassica napus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (82)...(1107)  
 US-10-223-076-6

## Alignment Scores:

Pred. No.: 2,2e-195 Length: 1446  
 Score: 1833.00 Matches: 342  
 Percent Similarity: 95.65% Conservative: 10  
 Best Local Similarity: 92.93% Mismatches: 16  
 Query Match: 66.15% Indels: 1  
 DB: 14 Gaps: 0

US-09-623-514A-2 (1-520) x US-10-223-076-6 (1-1446)

QY 153 GluAsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSer 172  
 DB 2 GAAATCTCATGAAGTACGGTGGTGTGATCAGACTGATTTCTGGTTTGTAGTTCAACGTCG 61  
 QY 173 LeuArgAspTrpProLeuPheMetCysIleSerLeuSerIlePheProLeuAlaIa 192  
 DB 62 CTGCGAGATG-CGCGTTTTCATGTGTGTCTCTCCCTTCAATCTTCTTCTTGGTGGTCC 120  
 QY 193 PheThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValIlePheLeu 212  
 DB 121 TTTACCGTCGAAATAGTACTCAGAAATGATATCTGAACCTGTGTCTATCTTTCTT 180  
 QY 213 HisIleIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAsp 232  
 DB 181 CATGTTATTATCACCATCACCAGGCTGTGTATCCAGTCTATGTCACCTTAAGGTGTAT 240  
 QY 233 SerAlaPheLeuSerGlyValThrLeuMetLeuThrCysIleValTrpLeuLysLeu 252  
 DB 241 TCTGCTCTTATCAGGTGACACGTGTATGCTCTCTCACTTCATTCGATGTGTGGCTGAAGT 300  
 QY 253 ValSerTrpAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAspLysAla 272  
 DB 301 GTTCTTACGCTCATCTAATCTAGACATAGAACCTAGCTAGTAACTATCTGTAAGGCC 360  
 QY 273 AsnProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetValAlaPro 292  
 DB 361 AATCTGAAGTCTCTACTATGTAGCTTGAAGCTTGGCTTATTCATCTGCTGCTCC 420  
 QY 293 ThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTrpValAla 312  
 DB 421 ACATTGTGTATCAGCCAAAGTATCCAGCTTCTCCATGTATCCGGAAGGTTGGTGCT 480  
 QY 313 ArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleGluClnTyr 332  
 DB 481 GGTCAATTTGCAAACTGGTCTATATTCTACTGACTCATGGGATTTATAATAGCAATAT 540

QY 333 IleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyrAlaIle 352  
 DB 541 ATAAATCTATTGTTAGGAACCTCAAAGCATCTCTGAAAGGGACCTTCTATATGCTATT 600  
 QY 353 GluArgValLeuLysLeuSerValProAsnLeuTyrValTrpLeuCysMetPheTyrCys 372  
 DB 601 GAAAGAGTGTGAAGCTTTTCAGTTCCAAATCTATATGTGTGGCTCTGCTATGCTTCTACTG 660  
 QY 373 PhePheHisLeuTyrLeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPhe 392  
 DB 661 TTCCTCCACCTTGTGTTAAACATATTGGCAGAGCTCTCTCTGCTGGGGACCGTGAATTC 720  
 QY 393 TyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetPro 412  
 DB 721 TACAAAGATTGGTGAATGCAAAAGCGTTGGAGATTATTGGAGAAATGTGGAATATGCT 780  
 QY 413 ValHisLysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIleProLys 432  
 DB 781 GTTCACAAATGGATGGTTCGACATGTATACCTTCCGTGCTGCGCATCAGATACCAAA 840  
 QY 433 ThrLeuAlaIleIleIleAlaPheLeuValSerAlaValAlaPheHisGluLeuCysIleAla 452  
 DB 841 GTACCCGCCATTATCATTTCTTCTAGTCTCTGAGTCTTTCATGAGTTATGTCATCGCA 900  
 QY 453 ValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeu 472  
 DB 901 GTTCCTTGGCGCTCTCTCAATCTATGGGCTTTCATGGGAATATGTTTCAGTCCCTTTG 960  
 QY 473 ValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePhe 492  
 DB 961 GTCTTTATCACAAATCTTTTACAAGAAAGTTTGGCTCCATGTTGGGAAACATGATCTTT 1020  
 QY 493 TrpPheIlePheCysIlePheGlyClnProMetCysValLeuLeuTyrTyrHisAspLeu 512  
 DB 1021 GGTTCAGCTTCTTGCATTTTCGGACACCGATGTGGGCTTCTTTATTACCATGACCTG 1080  
 QY 513 MetAsnArgLysGlySerMetSer 520  
 DB 1081 ATGAACCGCAAAAGGATCCATGCTCC 1104  
 RESULT 9  
 ; Sequence 17, Application US/10223076  
 ; Publication No. US20030074695A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Farese, Robert V  
 ; APPLICANT: Cases, Sylvaine  
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
 ; FILE REFERENCE: UCAL-105CIP3  
 ; CURRENT APPLICATION NUMBER: US/10/223,076  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 10/040,315  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 09/339,472  
 ; PRIOR FILING DATE: 1999-06-23  
 ; PRIOR APPLICATION NUMBER: 60/107,771  
 ; PRIOR FILING DATE: 1998-11-09  
 ; PRIOR APPLICATION NUMBER: PCT/US98/17883  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: 09/103,754  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 17  
 ; LENGTH: 1572  
 ; TYPE: DNA  
 ; ORGANISM: Zea mays  
 US-10-223-076-17

Alignment Scores:  
 Pred. No.: 3.96e-133 Length: 1572

```

: CURRENT APPLICATION NUMBER: US/10/223,076
:
: CURRENT FILING DATE: 2001-10-29
:
: PRIOR APPLICATION NUMBER: 10/040,315
:
: PRIOR FILING DATE: 2001-10-29
:
: PRIOR APPLICATION NUMBER: 09/339,472
:
: PRIOR FILING DATE: 1999-06-23
:
: PRIOR APPLICATION NUMBER: 60/107,771
:
: PRIOR FILING DATE: 1998-11-09
:
: PRIOR APPLICATION NUMBER: PCT/US98/17893
:
: PRIOR FILING DATE: 1998-08-28
:
: PRIOR APPLICATION NUMBER: 09/103,754
:
: PRIOR FILING DATE: 1998-06-24
:
: NUMBER OF SEQ ID NOS: 17
:
: SOFTWARE: FastSEQ for Windows Version 4.0
:
: SEQ ID NO 12
:
: LENGTH: 2099
:
: TYPE: DNA
:
: ORGANISM: Nicotiana tabacum
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: (208)...(1806)
:
: US-10-223-076-12

```

Alignment Scores:	3.48e-196	Length:	2099
Pred. No.:	1842.50	Matches:	361
Score:	76.82%	Conservative:	50
Percent Similarity:	76.82%	Mismatches:	100
Best Local Similarity:	76.48%	Indels:	25
Query Match:	66.49%	Gaps:	7
DB:	14		2

US-09-623-514A-2 (1-520) x US-10-223-076-12 (1-2099)

QY	20	1	MetaLalLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGlyGluPhe	20
Db	217	ATGGAAATCCGGAGAGCGTCAAAATGACGACGACGACGACGATTCGGGTATTGAGAAC	276	
QY	21	ValAspLeuAsp-----ArgLeuArgArgLysSerArgSerAspSerAsn	37	
Db	277	CTCACTCCGATCTTAATCACTCGGTTCGGAGGAGACGT-----GGCAGTAAT	324	
QY	38	GlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAla	57	
Db	325	GGTTTTGAGCGGCTAGTGCAATTAACAGTTCG-----GATCGCATATGACGGAA	375	
QY	58	AspValArgAspArgIleAspSer-----ValValAsnAspAspAlaGln	72	
Db	376	GATAGAGAGATGCTGTGTGGCAGCGGTGCTGGATTGAAACGGTGAATGACGCGGATAAA	435	
QY	73	GlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGly-AspAsnAsnGlyGlyGlyAr	92	
Db	436	TCGGTTGGTCACTCAGTGATGT-AATTCGAAAGGAGGACGACAGGAAATGATTAATGTGTC	494	
QY	92	gGlyGlyGlyGluArgGlyAsnAlaAspAlaThr-----PheThrTyrr	108	
Db	495	GAATGCTGAGGAAGCAAAATCAACGGAAACAACAACGACGCCGTTTAAATTCCTTCACAG	554	
QY	108	gProSerValProAlaHisArgArgAlaArgGluSerProLeuSerSerAspAlaIlePh	128	
Db	555	GCGCTCGCACCCAGCTCACCGGGCAAAATCAAGGAGAGTCCTCTCAGCTCCGACGCCATTTT	614	
QY	128	eLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValValLeuIleAlaValAsnSe	148	
Db	615	CAACACAGATCACCGACGCCCTGTTCATCTCTGTGTGTGTGCTGATGCTGCTTTAACAG	674	
QY	148	rArgLeuIleIleGluAsnLeuMetLysTyrrGlyTrpLeuIleArgThrAspPheTrpPh	168	
Db	675	CAGGCTGATTATCAGAGAATTGATGAAGTAGTGCCCTTTTAATTAGGCTGGCTTTTGGTT	734	
QY	168	eSerSerArgSerLeuArgAspTrpProLeuPheMetCysIleSerLeuSerIlePh	188	
Db	735	TAGCTCGAAGTCTGTTGAGGATTCGCGCGCTCTTAATGTGCTGCTCAGTCTCCAAATTTT	794	

## RESULT 8

US-10-223-076-6  
; Sequence '6, Application US/10223076  
; Publication No. US20030074695A1  
; GENERAL INFORMATION:

;; PRIOR FILING DATE: 1998-08-28  
;; PRIOR APPLICATION NUMBER: 09/103,754  
;; PRIOR FILING DATE: 1998-06-24  
;; NUMBER OF SEQ ID NOS: 17  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 14  
;; LENGTH: 1964  
;; TYPE: DNA  
;; ORGANISM: Piralla frutescens  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (69)...(1673)  
US-10-223-076-14

## Alignment Scores:

Pred. No.: 1.59e-197 Length: 1964  
Score: 1854.00 Matches: 359  
Percent Similarity: 74.18% Conservative: 49  
Best Local Similarity: 65.27% Mismatches: 92  
Query Match: 66.91% Indels: 50  
DB: 14 Gaps: 9

US-09-623-514a-2 (1-520) x US-10-223-076-14 (1-1964)

Qy 1 MetalalleuaspSerAlaGlyValThrValThr-----GluAsnGly 16  
Db 69 ATGCGATCTTGGACCTCGCGGAGATCTTGACACAGCAGCTCGTCCAGTGGCGGACACGCG 128  
Qy 17 GlyGlyGluPheValaspLeuArgLeuArgArgGlySerArgSerArgSerSer 36  
Db 129 GCGGCACATCACACCATCTTT-----CGCGGAGACAAAGTGGCGCTCCGTTCCGCGCT 182  
Qy 37 AsnGlyLeuLeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaPro 56  
Db 183 -----CTTCTCGACTCGGATCCAACTCTCTG----- 209  
Qy 57 AlaaspValArgaspArgIleaspSerValValasnasp-----AspAlaGlnGlyThr 74  
Db 210 -----GAGGCAGAGAGCGCAATCAATGATTCGGAATAATGTTCCGAACGAC 254  
Qy 75 AlaAsnLeuAlaGlyAspAsnAsnGlyGly----- 84  
Db 255 GCTAATTGATCGAAATCTCCGCGCGGAGCGCGTGAATCCGAGACGAAACACGAG 314  
Qy 85 -----GlyAspAsnAsnGly-----GlyGlyArgGlyGlyGlu 96  
Db 315 AGTTATGTAAGGAGGAGGCGGAAAGTGAAGGAGAGTGAAGAACTAGTAATGGCAAC 374  
Qy 97 GlyArgGlyAsnAlaaspAlaThrPheThrTyArgProSerValProAlaHisArgArg 116  
Db 375 GGAACCTGATGTTATGCGCGCTCAAAATTCACATTCAGGCGCGCGCTGCTCACCGCAA 434  
Qy 117 AlaArgGluSerProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPhe 136  
Db 435 AATAAGAGAGTCTCTTAGCTCCGAGCCATCTTCAACAGAGCCATGAGGCGCTTC 494  
Qy 137 AsnLeuCysValValValLeuAlaValasnSerArgLeuLeuIleGluAsnLeuMet 156  
Db 495 AACCTTGTATAGGGTCTGTGTGCTGTAATAAGACAGACTAATAATGAGAAATTAATG 554  
Qy 157 LysTyArgTrpLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrp 176  
Db 555 AAGTATGGTGGCTGATCAAAATCAGGATTTTGGTTTATGTTCAACATCGCTTAGGATGG 614  
Qy 177 ProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaIlePheThrValGlu 196  
Db 615 CCACCTGTAAGTGTCTTAGTCTTCCAGTTTTCACCTCGCTTCATCTTGTGTCGAG 674  
Qy 197 LysLeuValLeuGlnLysTyTrpValIleSerGluProValIlePheLeuHisIleIle 216  
Db 675 AAGTGTGTAACAACTAATATATACCTGAGTGGTCCAGCTCTTCTTCATGTTACATC 734  
Qy 217 ThrMetThrGluValLeuTyProValTyProValTyValThrLeuArgCysAspSerAlaPheLeu 236

Db 735 ACAACAGTGGAAATCTTGTTCAGTTGTTGTCATCTTAGGTGATGTTCTGCTGTCTA 794  
Qy 237 SerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyAla 256  
Db 795 TCAGGTGTACGCTAATGCTCTTGTGCTGCACTATGTTGGTGAAGCTCGTTTCTACGCA 854  
Qy 257 HisThrSerTyArgPheSerLeuAlaAsnAlaAlaAspLys----- 271  
Db 855 CATACAAACTATGATTTGAGAGTACTTGCATAATCACTTGTATAGTGGGAGCTATGTC 914  
Qy 272 -----AlaAsnProGluValSerTyTrpValSerLeuLysSerLeuAlaTyPheMet 289  
Db 915 AGGTACTTGGAACTCGACTAGCTTATAGCTTAAAGACTTGAAGACTTGCATTCGGAAGGT 974  
Qy 290 ValAlaProThrLeuCysTyArgProSerTyProArgSerAlaCysIleArgLysGly 309  
Db 975 GTTGCTCTACATTTGTACCAGCAAGCTACCTCGAGCTTGCATTCGGAAGGT 1034  
Qy 310 TrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIle 329  
Db 1035 TGGTGTGAAGGCAACTAATTAAGCTGTAATATTCACAGACTCATGGGATTTATATA 1094  
Qy 330 GluGlnTyTrpLeuAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu 349  
Db 1095 GAACAGTACATAAACCGGATCGTTCAAAATCTCAACATCTCTGAAAGAAACCTTTA 1154  
Qy 350 TyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyValTrpLeuCysMet 369  
Db 1155 TAGCCATTTAGAGGCTTGAAGCTTCTGTTCCAAATTTATATGTTGGCTCTGATG 1214  
Qy 370 PheTyrcysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysPheGlyAsp 389  
Db 1215 TTTTATGTTTTCACCTCTGGCTAAATATATCTGCTGAACCTCTGCTTTGGGAC 1274  
Qy 390 ArgGluPheTyrcysAspTrpPheAsnAlaLysSerValGlyAspTyTrpArgMetTrp 409  
Db 1275 CGTGAGTTTATAGGATTTGTTGAATCGGAGGACAGTGGAGGAGTACTGGAGAATCTGG 1334  
Qy 410 AsnMetProValHisLysTrpMetValArgHisIleTyrcysLeuArgSerLys 429  
Db 1335 AATATGCTGTCCATAAATGATGGTTCGGCATATATATCTGCTTACAAAATGGA 1394  
Qy 430 IleProLysThrLeuAlaIleIleAlaPheLeuValSerAlaValPheHisGluLeu 449  
Db 1395 ATACCAAGATAGTGGCAGTTTGTATCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1454  
Qy 450 CysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGln 469  
Db 1455 TGGCTTGCAGTCCCTTGCCTCAAAATATTCAAGTTTGGCGCTTCTCGGCTATCATCTTCAG 1514  
Qy 470 ValProLeuValPheIleThrAsnTyrcysLeuGlnGluArgPhe---GlySerThrValGly 488  
Db 1515 GTTCTCTCTGTAATCGTGAATAATTAATTCGCAAGAAAGTTCAAAAATCAATGCTGGC 1574  
Qy 489 AsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTy 508  
Db 1575 AATATGATGTTCTGGTCTTCTCTGATCTTCTTGGTCAACCTATGTTGTTGCTGCTGCT 1634  
Qy 509 TyrHisAspLeuMetAsnArgLysGlySer 518  
Db 1635 TACCACGACTTGAATCGAAGCAAGCTAGT 1664

## RESULT 7

US-10-223-076-12  
; Sequence 12, Application US/10223076  
; Publication No. US20030074695A1  
; GENERAL INFORMATION:  
; APPLICANT: Farese, Robert V  
; APPLICANT: Cases, Sylvaine  
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
; TITLE OF INVENTION: Uses thereof  
; FILE REFERENCE: UCAL-105CIP3



NAME/KEY: CDS  
LOCATION: (1)...(1512)  
US-10-223-076-8

## Alignment Scores:

Pred. No.: 7,55e-249 Length: 1512  
Score: 2309,50 Matches: 442  
Percent Similarity: 88.10% Conservative: 17  
Best Local Similarity: 84.84% Mismatches: 43  
Query Match: 83.35% Indels: 19  
DB: 14 Gaps: 5

US-09-623-514R-2 (1-520) x US-10-223-076-8 (1-1512)

```
Qy 1 MetAlaLeuAspSerAlaGlyValThrThr---ValThrGluAsnGlyGlyGlu 19
Db 1 ATGGCGATTTGGATTCTGGAGGCGTCGCTACCGCGGAGGAGACGGC-----51

Qy 20 PheValAspLeuAspArgLeuArgArgArgLysSerArgSerAspSerAsnGlyLeu 39
Db 52 GTCGCGGATCTCGACAGGCTCCACCGTCGTAATTCGAGTTCCGATTCTTCCAACGGACTC 111

Qy 40 LeuLeuSerGlySerAspAsnSerProSerAspValGlyAlaProAlaAspVal 59
Db 112 CTC-----TCCGTACTTCCCGTCGCGAGCATGTGGAGCTGGCGGCGCGAA 159

Qy 60 ArgAspArgLeuAspSerValAlaAsnAspAlaGlnGlyThrAlaAsnLeuAlaGly 79
Db 160 AGGATCGGGTTGATTCCGCTGCCGAGGAGGAGCTCAGGGAACAGCGAATTTAGCT---216

Qy 80 AspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGly 99
Db 217 -----GGCGAGATGCCGAACTAGGGAATCCCGCGGAGGC-----252

Qy 100 AsnAlaAspAlaThrPheThrArgProSerValProAlaHisArgAlaArgGlu 119
Db 253 -----CATGTAAAGTTTACGTATCGACGTCGCTCCAGCTATCGGAGGACGAGGAG 306

Qy 120 SerProLeuSerSerAspAlaLeuPheLysGlnSerHisAlaGlyLeuPheAsnLeuCys 139
Db 307 AGTCCTCTCAGCTCCGACGCTATCTTCAACAAGACCATCAGGATTGTTCAACCTCTGT 366

Qy 140 ValValValLeuLeuAlaValAsnSerArgLeuLeuLeuLeuLeuLeuLeuLeu 159
Db 367 GTAGTTCTTGTGTTGTTGTTAAACAGTAGACTCATCATCGAAAAACCTCATGAAGTATGGT 426

Qy 160 TrpLeuLeuArgThrAspPheThrPheSerArgSerLeuArgAspTTPProLeuPhe 179
Db 427 TGGTTGATCAGACTGATTTTGGTTTGTAGTTTCTACATCCTTACGAGACTGGCCGCTTTC 486

Qy 180 MetCysCysLeuSerLeuPhePheProLeuAlaAlaPheThrValGluLysLeuVal 199
Db 487 ATGGTTGTTCTTTTCCATTTCCGTTTCCGTTTGGCTGCCCTTCCAGGTCGAGAAATGTA 546

Qy 200 LeuGlnLysThrLeuSerGluProValValLeuPheLeuHisLeuLeuLeuLeuLeu 219
Db 547 CTTCAAGAAATCATATCATGACCTGTTGCCATCATCTTCTATGTCATTAATCAACATGACA 606

Qy 220 GluValLeuThrProValThrValThrArgCysAspSerAlaPheLeuSerGlyVal 239
Db 607 GAGTCTGTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666

Qy 240 ThrLeuMetLeuLeuThrCysLeuValTrpLeuLysLeuValSerThrAlaHisThrSer 259
Db 667 ACCTGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726

Qy 260 TyrAspLeuArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyr 279
Db 727 TACGACATAAGAACCCCTGGCAATTCAGCTGATAAGGTCGATCCTGAAATCTCTCTACTAT 786

Qy 280 ValSerLeuLysSerLeuAlaThrPheMetValAlaProThrLeuCysTyrGlnProSer 299
Db 787 GTTAGCTTGAGAGCTTGGCGTATTTCATGGTTGCTCCACACACTGTTTATCAGCCGANGC 846
```

```
Qy 300 TyrProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuVal 319
Db 847 TATCCACGTTCTCCATGTATCCGGAAGGTTGGTGCTGCTCAACTTGCAAAACTGGTC 906

Qy 320 IlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsn 339
Db 907 ATATTCTACTGGACTCATGGGATTATATAGAGCAATATATAATCTCTATTGTTAGAAC 966

Qy 340 SerLysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSer 359
Db 967 TCAAGCATCTCTGAAAGGGGACCTTCTATATGCTATTGAAAGAGTGTGAAGCTTCA 1026

Qy 360 ValProAsnLeuThrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsn 379
Db 1027 GTTCCAAATCTATATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1086

Qy 380 IleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAla 399
Db 1087 ATATTGCGAGAGCTCTCTGCTTCCGGGACCGTGAATCTACAAAGATTTGGTGAATGCA 1146

Qy 400 LysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysLysTrpMetValArg 419
Db 1147 AAAAGCCTTGGAGATTATGGAGATTGGAATATGCTGTTCAAAATGGATGGTTCGA 1206

Qy 420 HisLeuTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAla 439
Db 1207 CATGTATCTTCCGTCGCTGCGCATCAAGATCAACAAAGTACCGCCCATTTATCATTTGCT 1266

Qy 440 PheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLys 459
Db 1267 TTCTAGTCTCTGCGAGCTTTTCATGAGTTATGATGATGATGATGATGATGATGATGATGAT 1326

Qy 460 LeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeu 479
Db 1327 CTATGGCTTTTCATGGGAATATGTTTCAGTCCCTTTGGTCTTTATATCAAACTTTTA 1386

Qy 480 GlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePhe 499
Db 1387 CAAGAAAGGTTTGGCTCCATGTTGGTGGAAACATGATCTTTGGTTCAGCTTCTTGGCATTTTC 1446

Qy 500 GlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMet 519
Db 1447 GGACAAACCGATGTGTGGGCTTCTTTATACCATGATGATGATGATGATGATGATGATGATGAT 1506

Qy 520 Ser 520
Db 1507 TCC 1509
```

## RESULT 5

US-10-223-076-10  
Sequence 10, Application US/10223076  
Publication No. US20030074695A1  
GENERAL INFORMATION:  
APPLICANT: Farese, Robert V  
APPLICANT: Cases, Sylvaine  
TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
TITLE OF INVENTION: Uses thereof  
FILE REFERENCE: UCAL-105CIP3  
CURRENT APPLICATION NUMBER: US/10/223,076  
CURRENT FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: 10/040,315  
PRIOR FILING DATE: 2001-10-29  
PRIOR APPLICATION NUMBER: 09/339,472  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/107,771  
PRIOR FILING DATE: 1998-11-09  
PRIOR APPLICATION NUMBER: PCT/US98/17883  
PRIOR FILING DATE: 1998-08-28  
PRIOR APPLICATION NUMBER: 09/103,754  
PRIOR FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: FastSEQ for Windows Version 4.0



; LOCATION: (2)...(1107)  
US-10-223-076-4

**Alignment Scores:**

Pred. No.:	1,56-253	Length:	1537
Score:	2351.50	Matches:	447
Percent Similarity:	89.23%	Conservative:	17
Best Local Similarity:	85.96%	Mismatches:	37
Query Match:	84.86%	Indels:	19
DB:	14	Gaps:	5

US-09-623-514A-2 (1-520) x US-10-223-076-4 (1-1537)

Qy	1	MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe	20
Db	22	ATGGAGATTTTGGATCTCTGGAGCGCTCACTATGCCGACGGAGAACGGTGT	72
Qy	21	ValAspLeuAspArgLeuArgArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu	40
Db	73	GCGCATCTCGATACCTTCGTACCGGAACCGAGATCGGATCTCTCCATGACACTCT	132
Qy	41	LeuSerGlySerAspAsnSerProSerAspValGlyAlaProAlaAspValArg	60
Db	133	-----CCTGATCCGTACTGTTCCGAT-----GCTGACGTGAGG	168
Qy	61	AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp	80
Db	169	GATCGGTTGATTACAGCTGTT---GAGGATACTCAAGGAAAGCCAAATTTGGCGGAGAA	225
Qy	81	AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGluGlyArgGlyAsn	100
Db	226	AACGAA-----ATTAGGGAATCCGCTGGAGAAGCGGGGGGAAAC	264
Qy	101	AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgArgAlaArgGluSer	120
Db	265	GTGGATGTAAAGTACACGCTATCGGCGGTGGTTCAGCTCATCGGAGGTGCGGGAGAT	324
Qy	121	ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal	140
Db	325	CCACTCAGCTCTGACGCCATCTTCAACACAGAGCCACTCTGGACTATTCAACCTGTCTGTA	384
Qy	141	ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyr	160
Db	385	GTAGTCTCTGTTGCTGTAAACAGTAGACTCATCATCGAAATCTCATGAAGTAGTCGTTGG	444
Qy	161	LeuIleArgThrAspPheTyrPheSerSerArgSerLeuArgAspTyrProLeuPheMet	180
Db	445	TTGATCAGAACCTGATTCTTGTTGTTAGTTCAACGTCTCTGGAGATTGGCCCCCTTCATG	504
Qy	181	CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu	200
Db	505	TGTTGTCTCTCCCTTTCAATCTTTCCTTGGCTGCCCTTACCGTCGAGAAATTAGTACTT	564
Qy	201	GlnLysTyrIleSerGluProValIlePheLeuHisIleIleIleThrMetThrGlu	220
Db	565	CAGAAATGCATATCTGAACTCTGTGTCACTATTTCTTATATTATATCACCATGACCCGAG	624
Qy	221	ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr	240
Db	625	GTCTGTATCCAGTCTATGTCACTTAAGGTGTGATTCGCCCTCTTATCAGGTGTCACG	684
Qy	241	LeuMetLeuLeuThrCysIleValTyrLeuLysLeuValSerTyrAlaHisThrSerTyr	260
Db	685	TTTGATGCTCTTCACCTGGCAITGTGTGGCTGAAGTTGGTTCTTTACGCTCATACTACTAT	744
Qy	261	AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyrVal	280
Db	745	GACATAAGAACCCCTAGCTAATCTCATGTATGAAGCCCAATCCTGAAGTCTCCTACTATGTT	804
Qy	281	SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr	300
Db	805	ACGTTGAAGAGCTGGCGTATTTCATGCTGCTCCACATTTGTTGTTATCAGCGGACGAT	864

## US-10-157-855-1

## Alignment Scores:

Pred. No.: 1.76e-300 Length: 1942  
 Score: 2771.00 Matches: 520  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 13 Gaps: 0

## US-09-623-514A-2 (1-520) x US-10-157-855-1 (1-1942)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20  
 DB 237 ATGGCGATTTTGGATTTCGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGATTC 296  
 QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40  
 DB 297 GTCGATCTTGATAGCTTCGTCGACGGNAATCGAGATCGGATTTCTTCAACGGACTTCTT 356  
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60  
 DB 357 CTCCTCTGGTTCGGATAATAATCTCCTTCGGATGATGTTGGAGCTCCGCGCACGTTAGG 416  
 QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80  
 DB 417 GATCGGATGTGATCCGTTGTTAAACGATGACGCTCAGGGAACAGCCAAATTTGGCCGGAGAT 476  
 QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyArgGlyAsn 100  
 DB 477 AATAACGCTGGTGGGATTAATAACGGTGGTGGAGAGCGCGGAGAGAGAGAGAGAGAGAG 536  
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120  
 DB 537 GCCGATGTACGTTTACGTATACGACCGTCGGTTCAGCTCATCGAGGCGCAGAGAGAGT 596  
 QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140  
 DB 597 CCACCTAGCTCCGACGCAATCTTCAACAGAGACCATCGCGGATTTATCAACCTCTGTGTA 656  
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyr 160  
 DB 657 GTAGTTCTATTGCTGTAAACAGTAGACTCATCATCGAATCTTATGAAGTATGGTTGG 716  
 QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTyrProLeuPheMet 180  
 DB 717 TTGATCAGAACGGATTTCTGTTTAGTTCAAGATCGCTGCGAGATTTGGCCGCTTTTCATG 776  
 QY 181 CysCysIleSerLeuSerIlePheProLeuAlaIlePheThrValGluLysLeuValLeu 200  
 DB 777 TGTGTATATCCCTTTCGATCTTTCTTTCGCTGCTTACGGTGTGAGAAATTTGGTACTT 836  
 QY 201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220  
 DB 837 CAGAAATACATATCAGAACCTTGTGTATCTTTCTTCATATATATACACCTGACAGAG 896  
 QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240  
 DB 897 GTTTTGTATCCAGTTTACGTCACCTTAAGGTGTATCTGCTTTTATCAGGTGTCACT 956  
 QY 241 LeuMetLeuLeuThrCysIleValThrLeuLysLeuValSerTyrAlaHisThrSerTyr 260  
 DB 957 TTGATGCTCCCTTCACCTTTCGATTTGTGGTAAAGTTGGTTTCTTATGCTCATACGTAT 1016  
 QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrVal 280  
 DB 1017 GACATAGATCCCTAGCCATATGACGTATAGCCCAATCTCGAGTCTCTACTACGTT 1076  
 QY 281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300  
 DB 1077 AGCTTGAAGACCTTGGCATATTTTCATGGTCGCTCCACATGTGTATCAGCCCAAGTTAT 1136  
 QY 301 ProArgSerAlaCysIleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuValIle 320

DB 1137 CCACGTTCTGCTATATACGGAAGGTTGGTGGCTGCTCAATTTGCAAACTGGTCATA 1196  
 QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340  
 DB 1197 TTCACCGGATTCATGGGATTTATATAGAACAAATATATAATCTTATGTGAGAACTCA 1256  
 QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360  
 DB 1257 AAGCATCCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTTCGAAGCTTTCAGTT 1316  
 QY 361 ProAsnLeuTyrValThrLeuCysMetPheTyrCysPhePheHisLeuThrLeuAsnIle 380  
 DB 1317 CCAAAATTATATGTGTGGCTCTGCTGCTTCTACTGCTTCTTCCACCTTTGGTTAAACATA 1376  
 QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTyrTyrPAsnAlaLys 400  
 DB 1377 TTGGCAGAGCTTCTGCTTCGGGATCGTGAATCTTACAAGATTTGGTGGAAATGCANA 1436  
 QY 401 SerValGlyAspTyrTyrArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420  
 DB 1437 AGTGTGGGAGATTACTGGAGAAATGTGGAATATGCTGTTTCAATAATGGATGGTTCGACAT 1496  
 QY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440  
 DB 1497 ATATACCTCCGCTTTCGCGCAGCAAGATACCAAGACACTCGCCATTATCATTTGCTTTC 1556  
 QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460  
 DB 1557 CTAGTCTCGCAGCTTTCATGAGCTATGATCGAGCTTCTTCTGCTGCTCTTCAAGCTA 1616  
 QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480  
 DB 1617 TGGGCTTTTCTTGGGATTTATCTTTCAGGTGCTTGGTCTTTCATCACAACACTATCTACAG 1676  
 QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly 500  
 DB 1677 GAAAGGTTTGGCTCAACGGTGGGAAACATGATCTTCTGGTTCATCTTCTGCTTTCGGA 1736  
 QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520  
 DB 1737 CAACCGATGTGTGCTTCTTTATACCACGACCTGATGAACCCGAAAGGATCGATGTCA 1796

## RESULT 3

US-10-223-076-4  
 ; Sequence 4, Application US/10223076  
 ; Publication No. US20030074695A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Faresse, Robert V  
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
 ; FILE OF INVENTION: Uses Thereof  
 ; FILE REFERENCE: UCAL-105CIP3  
 ; CURRENT APPLICATION NUMBER: US/10/223,076  
 ; CURRENT FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 10/040,315  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 09/339,472  
 ; PRIOR FILING DATE: 1999-06-23  
 ; PRIOR APPLICATION NUMBER: 60/107,771  
 ; PRIOR FILING DATE: 1998-11-09  
 ; PRIOR APPLICATION NUMBER: PCT/US98/17883  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: 09/103,754  
 ; PRIOR FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 1537  
 ; TYPE: DNA  
 ; ORGANISM: Brassica napus  
 ; FEATURE:  
 ; NAME/KEY: CDS

```

; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (139)...(1701)
US-10-223-076-2

Alignment Scores:
Pred. No.: 1.7e-300 Length: 1904
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-623-514A-2 (1-520) x US-10-223-076-2 (1-1904)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
Db ATGGCGATTTCGATTCCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTC 198

QY 21 ValAspLeuAspArgLeuArgArgGlySerArgSerAspSerSerAsnGlyLeuLeu 40
Db GTTCGATCTGTAGAGCTTCGTCGCGGAAATCGAGATCGGATCTCTTAAACGAGCTCTT 258

QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyValAlaProAlaAspValArg 60
Db CTCTCTGGTTCGGATAAATCTCTCTCGGATGATGTTGGAGCTCCCGCGACGCTTAGG 318

QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
Db GATCGGATGATTCGTTTAAACATGACGCTCAGGGAACAGCCAAATTTGGCCGGAGAT 378

QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGly 100
Db AATAACGGTGGTGGCGATAAATACGGTGGTGGAGAGCGCGGAGAGAGAGAGAAC 438

QY 101 AlaAspAlaThrPheThrTyrrgProSerValProAlaHisArgAlaArgGluSer 120
Db GCCGATGCTACGTTTACGCTATCGACCGTGGTTCGAGCTTCATCGGAGCGCGAGAGAGT 498

QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
Db CCACCTTAGCTCCGACGCAATCTTCAACAGACGATGCGCGGATATTCACCTCTGTGTA 558

QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrrgGly 160
Db GTAGTCTTATTGCTGTAACACAGTAGACTCATCATCGAAATCTTATGAAGTATGTTGG 618

QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
Db TTGATCAGAACGGATTTCTGGTTTATGTTCAAGATCGCTCGAGATTTGGCCGCTTTTCATG 678

QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
Db TTTTGGTATATCCCTTCGATCTTCCTTTGGCTGCTTTACCGGTGAGAAATTTGGTACTT 738

QY 201 GlnLysTyrrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
Db CAGAAATACATATCAGAACCTGTTGTCATCTTCTTCATATATATATCATCACCATGACAGAG 798

QY 221 ValLeuTyrrProValTyrrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
Db GTTTTGTATCCAGTTTACGTCACCCAAAGGTGTGATCTCTGCTTTTATCATCAGGTGCTACT 858

QY 241 LeuMetLeuThrCysIleValTrpLeuLysLeuValSerTyrrAlaHisThrSerTyrr 260
Db TTGATGCTCCCTCACCTGTGATTTGGTAAAGTTTGTTCATATGCTCATCTACTAGCTAT 918

QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrrVal 280
Db GACATAAGATCCCTAGCCCAATCCAGCTGATAGGCCAATCTCTGAGTCTCTACTACGTT 978

```

```

281 SerLeuLysSerLeuAlaTyrrPheMetValAlaProThrLeuCysTyrrGlnProSerTyrr 300
979 AGCTTGAAGAGCTTGGCATATTTTCATGTCGCTCCACATTTGTTATCAGCCAAGTAT 1038

301 ProArgSerAlaCysIleArgLysGlyTrpValAlaAlaArgGlnPheAlaLysLeuValIle 320
1039 CCACGTTTCGCATGATATACGGAAGGTTGGTGGCTGCTCAATTTGCAAACTGGTCA 1098

321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrrIleAsnProIleValArgAsnSer 340
1099 TTCACCGGATTCATGGATTTATATAGAACATATATAAACTCTATTCAGGAACTCA 1158

341 LysHisProLeuLysGlyAspLeuLeuTyrrAlaIleGluArgValLeuLysLeuSerVal 360
1159 AAGCATCTCTTGAAGGCGATCTCTATATGCTATTGAAAGAGTGTGGAAGCTTTCAGT 1218

361 ProAsnLeuTyrrValTrpLeuCysMetPheTyrrCysPhePheHisLeuTrpLeuAsnIle 380
1219 CCAAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTCCACCTTTGGTTAAACATA 1278

381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrrLysAspTrpIrrpAsnAlaLys 400
1279 TTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATTCACAAAGATTTGGTGAATGCAAAA 1338

401 SerValGlyAspTyrrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
1339 AGTGTGGGAGATTTACTGGAGAATATGGAATATGCTGTTTCAATAAATGGATGCTTCACAT 1398

421 IleTyrrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
1399 ATATACTTCCCTGCTTGGCAGCAGAACAGATACCAAGACACTGCCATATATCATGTGCTTC 1458

441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
1459 CTAGTCTCTGCAGTCTTTCATGAGCTATGATGATGCTGCTGCTGCTCTCTCAAGCTA 1518

461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrrLeuGln 480
1519 TGGGCTTTCTTGGGATATGTTTTCAGGTGCTGCTTTCATCACAACATCTCTACAG 1578

481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
1579 GAAAGTTTGGCTCAACGGTGGGGAACATGCTCTCTGCTTCTCTCTCTCTCTCTCTCGA 1638

501 GlnProMetCysValLeuLeuTyrrTyrrHisAspLeuMetAsnArgLysGlySerMetSer 520
1639 CAACCGATGTGTGCTTCTTCTTATTACACGACCTGATGACCGAAAGGATCGATGCTCA 1698

RESULT 2
US-10-157-855-1
; Sequence 1, Application US/10157855
; Publication No. US20020170091A1
; GENERAL INFORMATION:
; APPLICANT: Lasser, Michael W.
; APPLICANT: Ruzinsky, Diane M.
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE OF INVENTION: Acid Sequences
; FILE REFERENCE: 16516.158
; CURRENT APPLICATION NUMBER: US/10/157,855
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/326,203
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: August 31, 2003, 05:57:06 ; Search time 371 Seconds

(without alignments)  
3215.662 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTENGGEF.....QPMCVLLYHYHDLNMRKGSMS 520

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip  
-Q/cgn2\_1/USPTO.spool/US09623514/runat\_29082003\_152202\_190/app\_query.fasta\_1.711  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=rnnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62  
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09623514@cgn2\_1.1.333@runat\_29082003\_152202\_190  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications\_NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	DB ID	Description
1	2771	100.0	1904 14	US-10-223-076-2 Sequence 2, Appli

Sequence 1, Appli  
Sequence 4, Appli  
Sequence 8, Appli  
Sequence 10, Appli  
Sequence 14, Appli  
Sequence 12, Appli  
Sequence 17, Appli  
Sequence 16, Appli  
Sequence 2, Appli  
Sequence 9, Appli  
Sequence 3, Appli  
Sequence 9, Appli  
Sequence 15, Appli  
Sequence 16, Appli  
Sequence 1, Appli  
Sequence 12, Appli  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 7, Appli  
Sequence 5, Appli  
Sequence 6, Appli  
Sequence 14, Appli  
Sequence 192, Appli  
Sequence 1, Appli  
Sequence 4, Appli  
Sequence 1, Appli  
Sequence 79, Appli  
Sequence 29, Appli  
Sequence 58, Appli  
Sequence 209, Appli  
Sequence 746, Appli  
Sequence 15694, A  
Sequence 279, Appli  
Sequence 11, Appli  
Sequence 10, Appli  
Sequence 4664, Ap  
Sequence 587, Appli  
Sequence 10, Appli  
Sequence 3, Appli  
Sequence 5, Appli  
Sequence 39256, A  
Sequence 7, Appli  
Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-10-223-076-2  
; Sequence 2, Application US/10223076  
; Publication No. US20030074695A1  
; GENERAL INFORMATION:  
; APPLICANT: Farese, Robert V  
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
; TITLE OF INVENTION: Uses Thereof  
; FILE REFERENCE: UCAL-105CIP3  
; CURRENT APPLICATION NUMBER: US/10/223.076  
; CURRENT FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 10/040,315  
; PRIOR FILING DATE: 2001-10-29  
; PRIOR APPLICATION NUMBER: 09/339,472  
; PRIOR FILING DATE: 1999-06-23  
; PRIOR APPLICATION NUMBER: 60/107,771  
; PRIOR FILING DATE: 1998-11-09  
; PRIOR APPLICATION NUMBER: PCT/US98/17883  
; PRIOR FILING DATE: 1998-08-28  
; PRIOR APPLICATION NUMBER: 09/103,754  
; NUMBER OF SEQ ID NOS: 17  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2

use as 10262

no Seq  
is not  
the same  
as Seq 2  
from

**THIS PAGE BLANK (uspto)**

Qy 423 FPCURSKIPKTLAIIIAFLVSAVFHELICIAVPCRLFKLWAFGLMFQVPLV-FTNTYLQE 481  
 Db 391 KPMRLRGSSKWMARTGVFLASAFHEYLVSPLRMFLWFTGMAQIPLAWFVGFQ 450  
 Qy 482 RFGSTVGNMIFWFIFCFIFGQPMCVLLYYHD 511  
 Db 451 NY-----GNAAVW-LSLIIGQPIAVLMYVHD 475

## RESULT 15

AAB19742  
 ID AAB19742 standard; Protein; 500 AA.

XX AC AAB19742;

XX DT 19-FEB-2001 (first entry)

XX DE Rat acyl CoA:cholesterol acyltransferase.

XX KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;  
 KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;  
 XX hypolipemic; rat.

XX OS Rattus sp.

XX PN W0200061771-A2.

XX PD 19-OCT-2000.

XX PF 12-APR-2000; 2000WO-US09696.

XX PR 12-APR-1999; 99US-0128995.

XX PA (MONS ) MONSANTO CO.

XX PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;  
 PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;

XX DR WPI; 2000-665136/64.

XX N-PSDB; AAA88846.

XX PT Genetically engineering the biosynthetic pathways in plants involved in  
 PT the accumulation of sterol compounds and tocopherol to produce  
 PT compounds for lowering the level of low density lipoprotein cholesterol  
 PT in blood serum

XX PS Disclosure; Page 64; 166pp; English.

XX CC The present sequence is that of rat acyl CoA:cholesterol  
 CC acyltransferase (ACAT), as deduced from ACAT DNA (see AAA88846).  
 CC Sterol O-acyltransferases such as ACAT catalyse the formation of  
 CC cholesterol esters from cholesterol and long chain fatty acids.  
 CC Recombinant constructs of the invention are used to alter the  
 CC biosynthesis and accumulation of sterols and tocopherols in  
 CC transgenic plants. Seeds of such plants may contain elevated  
 CC levels of sitostanol and/or its esters, and alpha-tocopherol, and  
 CC reduced levels of campesterol and campestanol and their esters.  
 CC The seeds may also contain the novel sterol brassicatanol. Oil  
 CC obtained from the seeds can be used in food and pharmaceutical  
 CC compositions to lower levels of low density lipoprotein cholesterol  
 CC in blood serum. ACAT enzymes can be used in the present invention  
 CC to produce elevated levels of phytosterol and/or phytostanol esters.

XX SQ Sequence 500 AA;

Query Match 27.9%; Score 773.5; DB 21; Length 500;

Best Local Similarity 37.5%; Pred. No. 1.4e-69;

Matches 194; Conservative 86; Mismatches 167; Indels 71; Gaps 18;

Qy 14 ENGSGEVDLRLRRRSKSDSNGL-LLSGSDNNSDDVAPADVRDRIDSVDNDAQ 72

Db 22 QGGSGPMVDEEVR-----DAAVGPDLAGGDAPAPA-PVPAPATRDK-----DRQ 67

Qy 73 GTANLAGDNNGGDNGGGGGGREGGRGNADATFTYRPSVPAHRARERPLSSDAIFKQSH 132  
 Db 68 TSV-----GDGHWEIR-----CH-RLQDSLFSSDSGF-SNY 96  
 Qy 133 AGLENLCVVLIIVANSRLIITENMKYGLWINTDFWFSRSLRD---WPLFMCCISLSIFP 189  
 Db 97 RGIILNCVVMILILSNARLFLENLIKYLIV-DPIQVVSLEFLKDPYSWPAACLIASNIFI 155  
 Qy 190 LAAPTVEKLVLQKYISBPVIFLHIIITMTTEVLVPVVTLCRCDSAFLSGVTLMILT-CIV 248  
 Db 156 VATFOIEKRUSVGALTEQOMGLLHVNLATIIICFPAAVALLVESITPVGSLFALASYII 215  
 Qy 249 WLKLVSYAHTSY-----DIRSLANAADK-----ANPEVSY--YVSLKSLAYFWVAPTL 295  
 Db 216 FLKLFSDYRDVNLWCRRVRKAKAVSAGKVKYSGAAQNTVSPDNLTYRDLVYFFAPTL 275  
 Qy 296 YQPSYPSRACIRKGVARQFAKLVFTGFMGFIIEQVINIVRNKHLKGLDLYA--IE 353  
 Db 276 YELNFRSPRIRKRLRRVLEMLFFTQLQVGLIQQWMPVPTIONSMPFK-DMDYSRIE 334  
 Qy 354 RVLKLSVNLVYVWLCMEYCFPHLWNLILAEILLCFGRDFYKDMWNAKSVGDYWRWNMPV 413  
 Db 335 RLLKLVPNHLIWLIFFTWLFHSCINAVAEILLQFGRDFYKDMWNAESVTFWQWNIPV 394  
 Qy 414 HKWMVRHIYFPCLRSKIPKTLAIIIAFLVSAVFHELICIAVPCRLFKLWAFGLMFQVPLV 473  
 Db 395 HKWCIRHFYKPMRLRSGSKWMARTGVFLASAFHEYLVSPLRMFLWFTGMAQIPLAWF 454  
 Qy 474 FTNLYQERFGSTVGNMIFWFIFCFIFGQPMCVLLYYHD 511  
 Db 455 WIVNRF---FQGNYGNAAVW-VTLIIGQPVAVLMYVHD 488

Search completed: August 31, 2003, 04:55:00

Job time : 73 secs

XX 02-DEC-1998; 98US-0110602.  
 PR 31-MAR-1999; 99US-0127111.  
 XX (DUPO ) DU PONT DE NEMOURS & CO E I.  
 XX Cahoon EB, Kinney AJ, Cahoon RE;  
 PI PI  
 XX WPI; 2000-412308/35.  
 DR N-PSDB; AAA48936.  
 XX Polynucleotides encoding diacylglycerol acetyltransferase, useful for  
 PT synthesis of triacylglycerols and increasing the level of oils in plant  
 PT seeds  
 XX Claim 12; Page 47-48; 62pp; English.  
 XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,  
 CC soybean and wheat were screened for sequences with homology to a  
 CC putative acyl CoA cholesterol acyltransferase related gene from  
 CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo  
 CC sapiens and Mus musculus. The cDNA clones identified from this process  
 CC were used to form complete diacylglycerol acyltransferase cDNA  
 CC sequences. The present sequence is corn diacylglycerol  
 CC acyltransferase protein, derived from a contig of clones  
 CC p0042.cspaf49r, p0122.ckmab57r and p0125.czaau61rb. Diacylglycerol  
 CC acyltransferases are involved in the synthesis of triacylglycerols.  
 CC Alteration of the expression of the diacylglycerol acyltransferase  
 CC DNA can be useful for increasing the level of oils in plant seeds.  
 CC Inhibitors of diacylglycerol acyltransferase may be useful as  
 CC herbicides.  
 XX Sequence 285 AA;  
 SQ  
 Query Match 37.2%; Score 1030.5; DB 21; Length 285;  
 Best Local Similarity 68.3%; Pred. No. 4.7e-96;  
 Matches 185; Conservative 30; Mismatches 47; Indels 9; Gaps 3;  
 QY 253 VSYATSDIRSLANADK-----ANPEVSYVSLKSLAYFMVAPILCYOPSPRSAC 305  
 DB 11 VSYATNDYDIRLSKSTEGKGAAYGVDPENMKDPTFKSLVYFMALPTLCYQTPQTC 70  
 QY 305 IRKGVARQFAKLVFTGFMGFIIOYINPIVRNSKHPKLDLLVAIERVLKSLVPLXV 365  
 DB 71 IRKGVWTOOLIKCVFTGLMGFIIOYINPIVRNSKHPKLDLLVAIERVLKSLVPLXV 130  
 QY 366 WLCMFYCFHMLNLAELLCGDFEYKDMWNKASVGDYWRMNNPVHKWVRHIYFPC 425  
 DB 131 WLCMFYCFHMLNLAELLCGDFEYKDMWNKASVGDYWRMNNPVHKWVRHIYFPC 190  
 QY 426 LRSKIPKTLAIIAFVSAVHEICIAVPCRLFKLWAFGLIMFOVPLVFTITNQLQERGS 485  
 DB 191 IRXGSRGVALIDISLVSNAVFHEICIAVPCRLFKLWAFGLIMFOVPLVFTITNQLQERGS 250  
 QY 486 T-VGNMFIWFIFICQPMCVLLYYHDLNMR 515  
 DB 251 VMVGNMFIWF-PSIVRQPMKLYNKHDMQK 280  
 RESULT 14  
 AAB15200  
 ID AAB15200 standard; protein; 488 AA.  
 XX  
 AC AAB15200;  
 XX  
 XX 14-DEC-2000 (first entry)  
 DT Human ACAT Related Gene Product 1 ARGPI.  
 XX  
 DE Human; ACAT Related Gene Product 1; ARGPI; gene therapy;  
 KW enzyme; acyl Coenzyme A-cholesterol acyltransferase 1; ACAT1;  
 KW sterol esterification; lipid homeostasis; diacylglycerol acyltransferase;  
 KW DGAT; cholesterol; triglyceride biosynthesis; hypertriglyceridaemia;

KW hyperlipidaemia; atherosclerosis; heart disease; obesity.  
 XX Homo sapiens.  
 XX US6100077-A.  
 XX 08-AUG-2000.  
 XX 01-OCT-1998; 98US-0165042.  
 XX 01-OCT-1998; 98US-0165042.  
 XX (UYCO ) UNIV COLUMBIA NEW YORK.  
 XX Sturley SL, Oelkers P;  
 XX WPI; 2000-557622/51.  
 XX N-PSDB; AAA76169.  
 XX New nucleic acid encoding a human diacylglycerol acyltransferase,  
 PT useful for treating hyperlipidaemia, atherosclerosis, heart disease, or  
 PT other diseases associated with an imbalance of triglyceride levels.  
 XX Claim 1; Fig 1A; 32pp; English.  
 XX The enzyme acyl Coenzyme A-cholesterol acyltransferase 1 (ACAT1) mediates  
 CC sterol esterification, an important component of intracellular lipid  
 CC homeostasis. The present sequence is human ACAT Related Gene Product 1  
 CC (ARGPI). This enzyme is a diacylglycerol acyltransferase (DGAT). This  
 CC enzyme does not esterify cholesterol. It is thought therefore that ARGPI  
 CC participates in the Coenzyme A-dependent acylation of substrate(s) other  
 CC than cholesterol e.g. diacylglycerol. Also, ARGPI has a predicted  
 CC diacylglycerol binding motif, suggesting that it may perform the last  
 CC acylation in triglyceride biosynthesis. ARGPI gene and protein are useful  
 CC for treating a subject who has an imbalance in triglyceride levels due to  
 CC a defect in esterification of diglycerol, via gene therapy. Particularly,  
 CC ARGPI is useful for treating hypertriglyceridaemia, hyperlipidaemia,  
 CC atherosclerosis, heart disease, obesity or other diseases associated with  
 CC high or excessive levels of triglyceride.  
 XX Sequence 488 AA;  
 SQ  
 Query Match 28.6%; Score 791.5; DB 21; Length 488;  
 Best Local Similarity 38.0%; Pred. No. 2e-71;  
 Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;  
 QY 25 RLRRKRSKSDSNGLLSGSDNNPSDDYDGPADYDRIDSDYVNDDAQ--TANLADNNG 83  
 DB 8 RRRRTGSRPSSHG-----GGPAAAE-----VRDAAGPDVGAAGDAPA 48  
 QY 84 GDDNNGGCGGEGEGEGRNADATYTPSPVPAHRRARESPSSDAIFKQSHAGLNFCLVVL 143  
 DB 49 PAPNKDGDAGVSGSHWELRC-----HRLQDSLFSDDSGF--SNYRGTLNMCVVM 96  
 QY 144 IAVNSRLIENLMKYGLIRTDWFSSRLRD---WPLFMCCISLIPPLAFTVEKVL 200  
 DB 97 ILSNARLFLENLIKIGILV-DPIQVVSLEKDPHWPAPCLVIAANVFAVAFQVEKRLA 155  
 QY 201 OKYISEPVVIFLHIITMTTEVLYPVYVTLRCDLSAFLSGVTLMLLT--CIWKLKLSVAHRS 259  
 DB 156 VGLTEQAGLLHLVLANLATILCFPAVLLVSEIIPVGSLLALMAHTILFLFSYR--- 212  
 QY 260 YDIRS-----LANAADKANP-----EVSY--YVSLKSLAYFMVAPILCYOPSPRSA 304  
 DB 213 -DVNSWCRRAKAAASAGKASSAAAPHVTSYDPNLTLYRDLYFFFAPTLCYELNFPSP 271  
 QY 305 CIRKGVARQFAKLVFTGFMGFIIOYINPIVRNSKHPKLDLLVA--IERVLKSLVNP 362  
 DB 272 RIRKRELLRILEMLFFTLQVGLTQQWNPQIOMKPFK--DMYSRIRIELRLKLAVEN 330  
 QY 363 LYVWLCMFYCFHMLNLAELLCGDFEYKDMWNKASVGDYWRMNNPVHKWVRHIY 422  
 DB 331 HLILWILFFWFLHSCNLNAVAELMQGDFEYKDMWNSESVTYFWQNNWNPVHKWIRH 390

```
XX SQ Sequence 361 AA;
Query Match 43.7%; Score 1211; DB 21; Length 361;
Best Local Similarity 61.18; Pred. No. 2.8e-114;
Matches 218; Conservative 49; Mismatches 80; Indels 10; Gaps 4;
QY 168 FSSRLRDWPLMCCISLIPPLAAFTVEKLVQKYSIEPVVIFLHIITMTEVLYPVV 227
DB 1 FNATSLRDWPLLMCCLSLPIPLGAFAVEKLAENNLVSDPATTCTFHILFTTEIVYVPV 60
QY 228 TLRCDSAFSLGVTMLLTCTIVWKLVSVAHTSYDIRSLANAADKANPEVS-----YV 280
DB 61 ILKCDSAVLSGFLVMTACIWLKLVSFATHTNDHDKLITSGKKVDNELTAAGIDNLQXP 120
QY 281 SLKSLAYFMVAPTLCYQPSYPRACIRKGVARQFAKLVFTGFMGFIIEQYINPIVRN 339
DB 121 TLGSLTFKMAPTLCYQAKVILRPYVRKGLVQRVLYLIFTGLQGFIIIEQYINPIVRN 180
QY 340 SKHPLKGLDLYAIERVLKSLVNPVNLVWLCMFYCFHLMNLILAEILCFGDREYKDWNA 399
DB 181 SOHPLMGGLLNAVETVLKSLPNVYLWLCMFYCLFHLNLILAEILRFGDREYKDWNA 240
QY 400 KSVGDYWRMNMVPHKVMVRIHYPPCLRSKIPKTLAIILAFVSAVPHELCIAVPCRLFK 459
DB 241 KTIDEYWRKWNMPVHKVIRVHIYPPCMRNGISKEVAVFISFFVSATLHEY-VLLFLHLK 299
QY 460 LWAFGLIMFQVPLVFTINTYQERFGST-VGNMIFWIFCFIGQPMCVLLYYHDLNMR 515
DB 300 FNAFLGMLQIPLIILSYLNKFSWTWGNMIFWFFCYIGQPMCVLLYYHDVNNR 356
RESULT 12
AAY94515
ID AAY94515 standard; Protein; 327 AA.
XX AC AAY94515;
XX DT 06-DEC-2000 (first entry)
XX DE Corn diacylglycerol acyltransferase protein #3.
XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
XX KW triacylglycerol; herbicide; EC2.3.1.20.
XX OS zea mays.
XX PN WO200032756-A2.
XX PD 08-JUN-2000.
XX PF 01-DEC-1999; 99WO-US28354.
XX PR 02-DEC-1998; 98US-0110602.
XX PR 31-MAR-1999; 99US-0127111.
XX (DUPO ) DU PONT DE NEMOURS & CO E I.
XX CAHoon EB, Kinney AJ, Cahoon RE;
XX WT: 2000-412308/35.
XX N-PSDB; AAA48935.
XX Polynucleotides encoding diacylglycerol acetyltransferase, useful for
XX synthesis of triacylglycerols and increasing the level of oils in plant
XX seeds
XX Claim 12; Fig 1; 62pp; English.
XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,
XX soybean and wheat were screened for sequences with homology to a
XX putative acyl CoA cholesterol acyltransferase related gene from
XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
```

```
CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is corn diacylglycerol acyltransferase
CC protein derived from clone cp1c.pk005.h23. Diacylglycerol
CC acyltransferases are involved in the synthesis of triacylglycerols.
CC Alteration of the expression of the diacylglycerol acyltransferase DNA
CC can be useful for increasing the level of oils in plant seeds.
CC Inhibitors of diacylglycerol acyltransferase may be useful as
CC herbicides.
XX SQ Sequence 327 AA;
Query Match 37.8%; Score 1046.5; DB 21; Length 327;
Best Local Similarity 56.0%; Pred. No. 1.3e-97;
Matches 191; Conservative 45; Mismatches 72; Indels 33; Gaps 4;
QY 166 FWFSSRLRDWPLMCCISLIPPLAAFTVEKLVQKYSIEPVVIFLHIITMTEVLYPV 225
DB 1 FWFNATSLRDWPLLMCCLSLPIPLGAFAVEKLAENNLVSDPATTCTFHILFTTEIVYVPV 60
QY 226 YVTLRCDSAFSLGVTMLLTCTIVWKLVSVAHTSYDIRSLANAADKANPEVS-----Y 278
DB 61 LVILKCDSAVLSGFLVMTACIWLKLVSFATHTNDHDKLITSGKKVDNELTAAGIDNLQ 120
QY 279 YVSLKSLAYFMVAPTLCYQPSYPRACIRKGVARQFAKLVFTGFMGFIIEQYINPIVR 338
DB 121 APTLGLSTYFMVAPTLCYQPSYPRTPYVRKGLVQRVLYLIFTGLQGFIIIEQYINPIVR 180
QY 339 NSKHLKGLDLYAIERVLKSLVNPVNLVWLCMFYCFHLMNLILAEILCFGDREYKDWNA 398
DB 181 NSQHPLMGGLLNAVETVLKSLPNVYLWLCMFYCLFHLNLILAEILRFGDREYKDWNA 240
QY 399 AKSVGDYWRMNMVPHKVMVRIHYPPCLRSKIPKTLAIILAFVSAVPHELCIAVPCRLF 458
DB 241 AKTIDEYWRKWNMPVHKVIRVHIYPPCMRNGISKEVAVFISFFVSATLHEVY----- 293
QY 459 KLWAFGLIMFQVPLVFTINTYQERFGSTVGNMIFWIFCFIF 499
DB 294 -----LLFHSSSAYI-NVI-----VLYFQMCPP 315
RESULT 13
AAY94516
ID AAY94516 standard; Protein; 285 AA.
XX AC AAY94516;
XX DT 06-DEC-2000 (first entry)
XX DE Corn diacylglycerol acyltransferase protein #4.
XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
XX KW triacylglycerol; herbicide; EC2.3.1.20.
XX OS zea mays.
XX FH Key Location/Qualifiers
XX FT Misc-difference 148 /note= "encoded by TAA"
XX FT Misc-difference 164 /note= "encoded by GNC"
XX FT Misc-difference 193 /note= "encoded by NAA"
XX FT Misc-difference 269 /note= "encoded by TNG"
XX FT Misc-difference 274 /note= "encoded by TAA"
XX PN WO200032756-A2.
XX PD 08-JUN-2000.
XX PF 01-DEC-1999; 99WO-US28354.
```



XX	DT	06-DEC-2000 (first entry)	XX
XX	DE	Wheat diacylglycerol acyltransferase protein #2.	XX
XX	KW	Diacylglycerol acyltransferase; corn; rice; soybean; wheat;	XX
XX	KW	triacylglycerol; herbicide; EC2.3.1.20.	XX
XX	OS	Triticum aestivum.	XX
XX	OS	WO200032756-A2.	XX
XX	PD	08-JUN-2000.	XX
XX	PD	01-DEC-1999; 99WO-US28354.	XX
XX	PR	02-DEC-1998; 98US-0110602.	XX
XX	PR	31-MAR-1999; 99US-0127111.	XX
XX	PA	(DUPO ) DU PONT DE NEMOURS & CO E I.	XX
XX	PI	Cahoon EB, Kinney AJ, Cahoon RE;	XX
XX	DR	WPI; 2000-412308/35.	XX
XX	DR	N-PSDB; AAA48942.	XX
XX	PT	Polynucleotides encoding diacylglycerol acetyltransferase, useful for	XX
XX	PT	synthesis of triacylglycerols and increasing the level of oils in plant	XX
XX	PT	seeds	XX
XX	PS	Claim 12; Fig 1; 62pp; English.	XX
XX	CC	In the present invention, cDNA libraries from Arabidopsis, corn, rice,	XX
XX	CC	soybean and wheat were screened for sequences with homology to a	XX
XX	CC	putative acyl CoA cholesterol acyltransferase related gene from	XX
XX	CC	Arabidopsis thaliana and diacylglycerol acyltransferases from Homo	XX
XX	CC	sapiens and Mus musculus. The cDNA clones identified from this process	XX
XX	CC	were used to form complete diacylglycerol acyltransferase cDNA	XX
XX	CC	sequences. The present sequence is wheat diacylglycerol	XX
XX	CC	acyltransferase protein, derived from clone wrl.pk0119.b6.fis.	XX
XX	CC	Diacylglycerol acyltransferases are involved in the synthesis of	XX
XX	CC	triacylglycerols. Alteration of the expression of the diacylglycerol	XX
XX	CC	acyltransferase DNA can be useful for increasing the level of oils in	XX
XX	CC	plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful	XX
XX	CC	as herbicides.	XX
XX	SS	Sequence 508 AA;	XX
XX	Query Match	57.6%; Score 1596; DB 21; Length 508;	XX
XX	Best Local Similarity	68.7%; Pred. No. 3.2e-153;	XX
XX	Matches 301; Conservative 44; Mismatches 81; Indels 12; Gaps 4;		XX
QY	88	NGGGRGGGEGRGADATFTYRFSV-PAHRRARESPSSDAIFKQSHAGLNCVYVLIAV 146	
DB	69	HGEAAGHAAAAARRDALL--PCVGAHRRVRKESPLSSDAIFRQSHAGLNCVYVLIAV 125	
QY	147	NSRLIITENLMKYGLLIRDFWSSRSRLRDWPLFMCCISLSIEPLAFTVEKVLQKYISE 206	
DB	126	NSRLIITENLMKYGLLIRAGFWFSARSGLDWPMLMCCLTPIPLAALMTEKWAQRKLND 185	
QY	207	PVYIFLHIITTEVLYPVVYVTLRCDASFSLGVTMLLMTCIVWLKLVSAHTSYDIRSLA 266	
DB	186	HVSILLHIITTVLYPVVYVILKCSAVLSGFVLMFIASITWLKLVSAHTNYDIRILS 245	
QY	267	NAADKA-----NPEVSYYVLSKLSIAYPMVAPTICYPSPRACIRKGWVARQFAKLV 319	
DB	246	QSTIEKGATHGSSIDEENIKGPTINSVYVFMPLAPTICYPSPRTAFIRKGWVTRQLIKCV 305	
QY	320	IFGCFWCFITEQVINTVNSKHPKGLDLYLAIETAVRLKLSVNLVYVLMCFYCFEHLWLN 379	
DB	306	VFTGLMGFITEQVINTVNSKHPKGLDLYLAIETAVRLKLSVNLVYVLMCFYCFEHLWLN 365	
QY	380	ILAEELFCFGDREYFKDWNNAKSYGDYWRMNNMPVHKWYVRHIYFFPCILRSKIPKTLAIIITA 439	

```

SQ Sequence 504 AA;
Query Match 65.7%; Score 1821; DB 21; Length 504;
Best Local Similarity 66.5%; Pred. No. 4.2e-176;
Matches 353; Conservative 47; Mismatches 91; Indels 40; Gaps 9;

QY 1 MAILDS-AGVTVTVENGGEFVDLRLRRKRSRSSNGLLLSGSDNNSPSDVGCAPADV 59
DQ 1 MAIDPEPSVATALHSS-----LRRRPS-ATSTAGLFNSPETTTDSSGDDDLAKDSG 51
QY 60 RDRIDSVDNDAOGTANLAGDNNGGDNGGGRGGEGGNADAT---FTVRPSVPAHRR 116
DQ 52 SD--DSINSDDA--AVNSQQONE-----KQDIDFSLAFATRPSVPAHRR 92
QY 117 ARESPSSDAIFKQSHAGLNLVAVNSRLIENLMKYGLWIRTDFFSSSLRDW 176
DQ 93 VKESPLSSDTIFRQSHAGLNLVAVNSRLIENLMKYGLWIKSGFESSSLRDW 152
QY 177 PLFMCISLSTPLAFTVEKLVQKYSIEPVVPLHIIITMTVEVLYPVVYVTLRCDAPL 236
DQ 153 PLFMCCLSLVFPFAFIVEKLAQRKCIPEPVVVLHIIITSTLSLFPVVLRLCDAPL 212
QY 237 SGVTLMLTCTVWKLVSVAHTSYDIRSLANAADKA-----NPEVSYVYSLKSLAYFM 289
DQ 213 SCVTLMLTSCVWKLVSVAHTSYDIRSLANAADKA-----NPEVSYVYSLKSLAYFM 272
QY 290 VAPTLCYQSPRSACIRKGVARQFAKLVITFTGFMGFIIEQYINPIVNSKHPKLGDL 349
DQ 273 VAPTLCYQSPRTPYIRKGLVRLVLIIFTGVMGFIIDQYINPIVNSKHPKLGDL 332
QY 350 YAIERVLKSLVPLNLYWLCMYCFPHLMNLIAELLCFGRDFEYKDMWNAKSVGDYWRMW 409
DQ 333 YATERVLKSLVPLNLYWLCMYCFPHLMNLIAELLCFGRDFEYKDMWNAKSVGDYWRMW 392
QY 410 NMPVHKWVRHYFPCLRSKIPKTLAIIIAFLVSFAVHELCLAVPCRLFKLWAFGIMFO 469
DQ 393 NMPVHKWVRHYFPCLRSKIPKTLAIIIAFLVSFAVHELCLAVPCRLFKLWAFGIMFO 452
QY 470 VPLVFITNLYQERF-GSTVGNMIFWIFCIFGQPMCVLLYVHDLNMRKGM 519
DQ 453 VPLVLTITNLYQERF-GSTVGNMIFWIFCIFGQPMCVLLYVHDLNMRKGM 503

RESULT 9
AA94518
ID AA94518 standard; Protein; 500 AA.
XX AC AA94518;
XX DT 06-DEC-2000 (first entry)
XX DE Rice diacylglycerol acyltransferase protein #2.
XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
XX KW triacylglycerol; herbicide; EC2.3.1.20.
XX OS Oryza sativa.
XX PH Key Location/Qualifiers
XX FT 1. 69
XX FT /note= "The nucleotides encoding this region are
XX FT not given in AAA48938"
XX FT 70..500
XX FT /note= "Encoded by nucleotides 15 to 1310 of the
XX FT cDNA in AAA48938"
XX PN WO200032756-A2.
XX PD 08-JUN-2000.
XX PF 01-DEC-1999; 99WO-US28354.
XX PF 02-DEC-1998; 98US-0110602.
PR

```

```

PR 31-MAR-1999; 99US-0127111.
XX PA (DUPO ) DU PONT DE NEMOURS & CO E I.
XX PI Cahoon EB, Kinney AJ, Cahoon RE;
XX WPI; 2000-412308/35.
DR N-PSDB; AAA48938.
XX PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
XX synthesis of triacylglycerols and increasing the level of oils in plant
XX seeds
XX Claim 12; Fig 1; 62pp; English.
XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,
XX soybean and wheat were screened for sequences with homology to a
XX putative acyl CoA cholesterol acyltransferase related gene from
XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
XX sapiens and Mus musculus. The cDNA clones identified from this process
XX were used to form complete diacylglycerol acyltransferase cDNA
XX sequences. The present sequence is rice diacylglycerol
XX acyltransferase protein, derived from clone rl324.pk0034.d8.fis.
XX Diacylglycerol acyltransferases are involved in the synthesis of
XX triacylglycerols. Alteration of the expression of the diacylglycerol
XX acyltransferase DNA can be useful for increasing the level of oils in
XX plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful
XX as herbicides.
XX Sequence 500 AA;
Query Match 59.1%; Score 1636.5; DB 21; Length 500;
Best Local Similarity 65.6%; Pred. No. 2.4e-157;
Matches 292; Conservative 61; Mismatches 79; Indels 13; Gaps 3;

QY 84 GGDNG-----GGGGGGGGRGNADATFTYRSPVPAHRRARSPSSDAIFKQSHAGLNL 138
DQ 51 GGDNSGRVLRPGGGGGGGGDFSAFTFRAAAPVHRKAKESPLSSDAIFKQSHAGLNL 110
QY 139 CVVLIIVANSRLIENLMKYGLWIRTDFFSSRLDWPFLPMCCISLSTPLAFTVEKL 198
DQ 111 CIVLVAVNSRLIENLMKYGLLIRAGFWFNDKSLRDWPLLMCCLSLPAFPLGAFVAKL 170
QY 199 VLQYIIEPVVIFLHIIITMTVEVLYPVVYVTLRCDAPLGLVTLMLTCTVWKLVSVAHT 258
DQ 171 AFNVITDAVATCLHIFLSTTEIVYVVLVKCDSAVLSCGFLLIIFACIWLKVSFAHT 230
QY 259 SYDIRSLANAADKANPEVSY-----VSLKSLAYFVAVPTLCYQSPRSACIRKGMV 311
DQ 231 NHDIRQLTMGGKKVDNELSTVDMNDLQPTTLGNLIYFMWAPTLCYQSPRTPSCYRKGWL 290
QY 312 ARQFAKLVITFTGFMGFIIEQYINPIVNSKHPKLGDLVYATIERVLKSLVPLNLYWLCMFY 371
DQ 291 IRQIILYLIFTGLQGFIEQYINPIVNSKHPKLGDLVYATIERVLKSLVPLNLYWLCMFY 350
QY 372 CFFHLWNLIALBLLCFGRDFEYKDMWNAKSVGDYWRMNMPVHKWVRHYFPCLRSKIP 431
DQ 351 AFFHLWLSILAEILRFGRDFEYKDMWNAKTIIDYWRKWNMPVHKWVRHYFPCLRNIGS 410
QY 432 KTLAIIIAFLVSFAVHELCLAVPCRLFKLWAFGIMFOVPLVITNLYQERFST-VGNM 490
DQ 411 KEVAVLISFLVSFAVHELCLAVPCRLFKLWAFGIMFOVPLVITNLYQERFST-VGNM 470
QY 491 IFWTFICIFGQPMCVLLYVHDLNMR 515
DQ 471 IFWTFICIFGQPMCVLLYVHDLNMR 495

RESULT 10
AA94522
ID AA94522 standard; Protein; 508 AA.
XX AC AA94522;

```

XX DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;  
 KW size; weight; carbon flux; TAGI; insertion mutant.  
 XX Arabidopsis thaliana.  
 OS  
 XX Key Location/Qualifiers  
 FH Misc-difference 158...184  
 FT /label= insertion  
 FT /note= "present due to a 147 bp insertion duplication  
 in the genomic DNA, see AAA51485".  
 XX  
 XX WO200036114-A1.  
 XX  
 XX 22-JUN-2000.  
 XX  
 XX 16-DEC-1999; 99WO-CA01202.  
 XX  
 XX 17-DEC-1998; 98US-0112812.  
 XX  
 XX (CANADA) NAT RES COUNCIL CANADA.  
 XX  
 XX Zou J, Taylor DC, Wei Y, Jaka CC;  
 PI WPI; 2000-431592/37.  
 DR N-PSDB; AAA51484, AAA51485.  
 XX  
 XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis  
 PT thaliana for transforming plants and regulating seed oil content, fatty  
 PT acid synthesis and seed oil acyl composition in commercial and crop  
 PT plants  
 XX  
 XX Claim 5; Page 81-83; 91pp; English.  
 XX  
 XX This is the Arabidopsis thaliana ecotype Columbia AS11 insertion mutant  
 CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a  
 CC 147 bp insertion located at the central region of intron 2. The insertion  
 CC is a duplication of a segment that is composed of 12 bp from the 3' end  
 CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the  
 CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful  
 CC for regulating seed oil content, the ratio of diacylglycerol to  
 CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil  
 CC acyl composition, seed size/weight and carbon flux into other seed  
 CC components in commercial and crop plants. The natural formation of  
 CC triacylglycerols can be modified to increase the yield in commercial  
 CC plant oils or modify their composition to achieve specific commercial  
 CC improvements of plants and plant products.  
 XX  
 XX Sequence 547 AA;  
 SQ  
 Query Match 99.2%; Score 2747.5; DB 21; Length 547;  
 Best Local Similarity 95.1%; Pred. No. 2.8e-270;  
 Matches 520; Conservative 0; Mismatches 0; Indels 27; Gaps 1;  
 QY 1 MAILDSAGTIVTENGGEFVDLRLRRKRSRSDSSNGLLLSGSDNNSPSDDVGAPADVR 60  
 DB 1 MAILDSAGTIVTENGGEFVDLRLRRKRSRSDSSNGLLLSGSDNNSPSDDVGAPADVR 60  
 QY 61 DRIDSVNDDAAGTANLAGDNNGGDNGGGRGGEGRGADATFTYRPSVPAHRRARES 120  
 DB 61 DRIDSVNDDAAGTANLAGDNNGGDNGGGRGGEGRGADATFTYRPSVPAHRRARES 120  
 QY 121 PLSSDAIFKQ-----SHAGFLNLCVVLVAVNSRLIIE 153  
 DB 121 PLSSDAIFKQ-----SHAGFLNLCVVLVAVNSRLIIE 180  
 QY 154 NLMYGLWIRLDFSSRLRDWFLMCCISLSIFPLAAFTVEKLVQKYSPEVWIFLH 213  
 DB 181 NLMYGLWIRLDFSSRLRDWFLMCCISLSIFPLAAFTVEKLVQKYSPEVWIFLH 240  
 QY 214 IIITWTEVLYPVVYTLRCDSAFLSGVTMLLTCTIWLKLVSYAHTSYDIRSLANAADKAN 273  
 DB 241 IIITWTEVLYPVVYTLRCDSAFLSGVTMLLTCTIWLKLVSYAHTSYDIRSLANAADKAN 300

QY 274 PEVSYYVSLKSLAYPMVAPTLCYQPSYRACIRKGVARQFAKLVIFTGFMGFIIEQYI 333  
 DB 301 PEVSYYVSLKSLAYPMVAPTLCYQPSYRACIRKGVARQFAKLVIFTGFMGFIIEQYI 360  
 QY 334 NPIVNSKHPLKGLDLYAIAERVLKLSVPLNLYWLCMFYCFPHLWLNILAEILLCFGDREFY 393  
 DB 361 NPIVNSKHPLKGLDLYAIAERVLKLSVPLNLYWLCMFYCFPHLWLNILAEILLCFGDREFY 420  
 QY 394 KDMWNAKSVGDIYWRMNMNPFVHKWVRHIYFFCLRSKIPKTLAIITAFVLSAVFHELCTAV 453  
 DB 421 KDMWNAKSVGDIYWRMNMNPFVHKWVRHIYFFCLRSKIPKTLAIITAFVLSAVFHELCTAV 480  
 QY 454 PCRLFKLWAFGLGIMFQVPLVFTITNYLQERFSGTVGNMIFWIFCITFGQPMCVLLYYHDLN 513  
 DB 481 PCRLFKLWAFGLGIMFQVPLVFTITNYLQERFSGTVGNMIFWIFCITFGQPMCVLLYYHDLN 540  
 QY 514 NRKGSMS 520  
 DB 541 NRKGSMS 547  
 RESULT 8  
 AAY94519  
 ID AAY94519 standard; Protein; 504 AA.  
 XX  
 AC AAY94519;  
 DT 06-DEC-2000 (first entry)  
 XX  
 DE Soybean diacylglycerol acyltransferase protein #1.  
 XX  
 KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;  
 triacylglycerol; herbicide; EC2.3.1.20.  
 OS Glycine max.  
 XX  
 PN WO200032756-A2.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 01-DEC-1999; 99WO-US28354.  
 XX  
 PR 02-DEC-1998; 98US-0110602.  
 PR 31-MAR-1999; 99US-0127111.  
 XX  
 PA (DUPO) DU PONT DE NEMOURS & CO E I.  
 XX  
 PI Cahoon EB, Kinney AJ, Cahoon RE;  
 XX  
 DR WPI; 2000-412308/35.  
 DR N-PSDB; AAA48939.  
 XX  
 PT Polynucleotides encoding diacylglycerol acyltransferase, useful for  
 synthesis of triacylglycerols and increasing the level of oils in plant  
 seeds  
 XX  
 PS Claim 4; Page 51; 62pp; English.  
 XX  
 CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,  
 CC soybean and wheat were screened for sequences with homology to a  
 CC putative acyl CoA cholesterol acyltransferase related gene from  
 CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo  
 CC sapiens and Mus musculus. The cDNA clones identified from this process  
 CC were used to form complete diacylglycerol acyltransferase cDNA  
 CC sequences. The present sequence is soybean diacylglycerol  
 CC acyltransferase protein, derived from clone srl.pk0098.a8.  
 CC Diacylglycerol acyltransferases are involved in the synthesis of  
 CC triacylglycerols. Alteration of the expression of the diacylglycerol  
 CC acyltransferase DNA can be useful for increasing the level of oils in  
 CC plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful  
 CC as herbicides.  
 XX

QY	Sequence	520 AA;
	Query Match 100.0%; Score 2771; DB 22; Length 520; Best Local Similarity 100.0%; Pred. No. 1.1e-272; Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MAILDSAGVTVTVTENGCGGEFVDLRLRRRKRSDSSNGLLLSGSDNNSPSDDVCGAPADVR 60	
Db	1 MAILDSAGVTVTVTENGCGGEFVDLRLRRRKRSDSSNGLLLSGSDNNSPSDDVCGAPADVR 60	
QY	61 DRIDSVVYNDDAQGTANLAGDNNGGDNGGRGGEGRGGNADATFTYRPSVPAHRRARES 120	
Db	61 DRIDSVVYNDDAQGTANLAGDNNGGDNGGRGGEGRGGNADATFTYRPSVPAHRRARES 120	
QY	121 PLSSDAIFKQSHAGLFNLCVYVLIANSRLIENLMKYGLIRTDFFWSSRLSDWPLFM 180	
Db	121 PLSSDAIFKQSHAGLFNLCVYVLIANSRLIENLMKYGLIRTDFFWSSRLSDWPLFM 180	
QY	181 CCISLSIFPLAAFTVEKLVQKYLISEPWVIFLHIIITMTVEVLYPVYVTLRCDSAFLSGVT 240	
Db	181 CCISLSIFPLAAFTVEKLVQKYLISEPWVIFLHIIITMTVEVLYPVYVTLRCDSAFLSGVT 240	
QY	241 LMLLTCIWLKLSYAHSTYDIIRSLANAADKANPEVSYVYSLKSLAYFMVAPTLCYOPSY 300	
Db	241 LMLLTCIWLKLSYAHSTYDIIRSLANAADKANPEVSYVYSLKSLAYFMVAPTLCYOPSY 300	
QY	301 PRSACIRKGWVARQFAKLVIFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIERVLKLSV 360	
Db	301 PRSACIRKGWVARQFAKLVIFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIERVLKLSV 360	
QY	361 PNLVYVWLCMYCFCFHLWNLITAEALLCGDREFYKDWNNAKSVGDYWRWNNMPVHKWVRH 420	
Db	361 PNLVYVWLCMYCFCFHLWNLITAEALLCGDREFYKDWNNAKSVGDYWRWNNMPVHKWVRH 420	
QY	421 IYPPCLRSKIPKTLAIIIAFLVSAVFHELCIATVPCRFLKFWAFLGIMFQVPLVITNVLQ 480	
Db	421 IYPPCLRSKIPKTLAIIIAFLVSAVFHELCIATVPCRFLKFWAFLGIMFQVPLVITNVLQ 480	
QY	481 ERGSGTVGNMIFWFIFCFIGOPMCVLLYYHDLNMRKGSMS 520	
Db	481 ERGSGTVGNMIFWFIFCFIGOPMCVLLYYHDLNMRKGSMS 520	

RESULT 6	
AA94512	
ID	RAY94512 standard; Protein; 520 AA.
XX	
XX	AA94512;
XX	
XX	
DT	06-DEC-2000 (first entry)
XX	
DE	Arabidopsis diacylglycerol acyltransferase clone araebF protein.
XX	
XX	Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW	triacylglycerol; herbicide; EC2.3.1.20.
KW	

Polynucleotides encoding diacylglycerol acyltransferase, useful for synthesis of triacylglycerols and increasing the level of oils in plant seeds -

Claim 14; Fig 1; 62pp; English.

In the present invention, cDNA libraries from Arabidopsis, corn, rice, soybean and wheat were screened for sequences with homology to a putative acyl CoA cholesterol acyltransferase related gene from Arabidopsis thaliana and diacylglycerol acyltransferases from Homo sapiens and Mus musculus. The cDNA clones identified from this process were used to form complete diacylglycerol acyltransferase cDNA sequences. The present sequence is Arabidopsis thaliana diacylglycerol acyltransferase protein, derived from clone araebcf. Diacylglycerol acyltransferases are involved in the synthesis of triacylglycerols. Alteration of the expression of the diacylglycerol acyltransferase DNA can be useful for increasing the level of oils in plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful as herbicides.

Sequence 520 AA;

Query Match 99.4%; Score 2753; DB 21; Length 520;  
Best Local Similarity 99.6%; Pred. No. 7.2e-271;  
Matches 518; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTENGSGGEFVDLRLRRRSRSDSSNGLLSGDNNSPDDVGAPADVR 60  
Db 1 MAILDSAGVTVTENGSGGEFVDLRLRRRSRSDSSNGLLSGDNNSPDDVGAPADVR 60  
QY 61 DRIDSVVNDDAQSTANLAGDNNGGDNGGCGGEGRGNADATFTYRPSVPAHRARES 120  
Db 61 DRIDSVVNDDAQSTANLAGDNNGGDNGGCGGEGRGNADATFTYRPSVPAHRARES 120  
QY 121 PLSDDAIFKQSHAGLFNLCVVVLIIVNSRLIIENLMKYGLIRTFDFWSSRLRDWPLFM 180  
Db 121 PLSDDAIFKQSHAGLFNLCVVVLIIVNSRLIIENLMKYGLIRTFDFWSSRLRDWPLFM 180  
QY 181 CCLSLSTFPLAAFTVEKLVLOKYSBPVVFLHIIITMTVEVLPVYVTLRCDSAFLSGVT 240  
Db 181 CWSLSLFFPLAAFTVEKLVLOKYSBPVGFILHIIITMTVEVLPVYVTLRCDSAFLSGVT 240  
QY 241 LMLLTCIVWLKLSYAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFWVAPLTCQPSY 300  
Db 241 LMLLTCIVWLKLSYAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFWVAPLTCQPSY 300  
QY 301 PRSACIRKGWVARQFAKLVITFGFMGFIIEQYINPIVRNSKHPKLGDLLYAIEVRLKLSV 360  
Db 301 PRSACIRKGWVARQFAKLVITFGFMGFIIEQYINPIVRNSKHPKLGDLLYAIEVRLKLSV 360  
QY 361 PNLVYVWLCMFCYCFHLLWNLTLABELCGDREFYKDWNAKSVGDIYWRMNMPVHKWVRH 420  
Db 361 PNLVYVWLCMFCYCFHLLWNLTLABELCGDREFYKDWNAKSVGDIYWRMNMPVHKWVRH 420  
QY 421 IYFPCLRSKIPKTLAIIIAFLVSAVFHELICIAVPCRLFKLWAEFLGIMQVPLFIITNYLQ 480  
Db 421 IYFPCLRSKIPKTLAIIIAFLVSAVFHELICIAVPCRLFKLWAEFLGIMQVPLFIITNYLQ 480  
QY 481 ERFGSTVGNMIFWFIICIFGQPMCVLLYYHDLNMRKGSMS 520  
Db 481 ERFGSTVGNMIFWFIICIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 7  
AA96854  
ID AA96854 standard; Protein; 547 AA.  
XX  
AC AA96854;  
XX  
XX 09-OCT-2000 (first entry)  
XX  
XX A. thaliana AS11 mutant diacylglycerol acyltransferase.

RESULT 7	
AA96854	
ID	AA96854 standard; Protein; 547 AA.
XX	
XX	AA96854;
XX	
XX	AC
XX	AC
XX	09-OCT-2000 (first entry)
XX	
XX	
DE	A. thaliana AS11 mutant diacylglycerol acyltransferase.

XX OS Arabidopsis thaliana.  
 XX PN WO9963096-A2.  
 XX PD 09-DEC-1999.  
 XX PF 04-JUN-1999; 99WO-US12541.  
 XX PR 05-JUN-1998; 98US-0088143.  
 XX PR 12-NOV-1998; 98US-0108389.  
 XX PA (CALJ ) CALGENE LLC.  
 XX PI Lassner MW, Ruezinsky DM;  
 XX DR WPI; 2000-105701/09.  
 XX DR N-PSDB; AAZ45371.  
 XX PT Novel polynucleotides used for modifying plant oil composition and for  
 XX PT developing products for treating e.g. cancer, diabetes, cardiopulmonary  
 XX PT disease or metabolic disorders -  
 XX PS Example 2; Page 65-67; 89pp; English.  
 XX CC The present sequence represents an acyl-CoA:cholesterol acyltransferase  
 XX CC (ACAT) related protein. The ACAT-like protein is active in the formation  
 XX CC of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and  
 XX CC sterol and/or diacylglycerol substrate. The DNA can be used for  
 XX CC modifying the lipid composition of plant cells. The ACAT-like protein  
 XX CC has diacylglycerol acyltransferase (DAGAT) activity, and so the  
 XX CC synthesis of triglycerides can be suppressed or increased using the  
 XX CC DNA. The protein can be used to produce plant oils with a modified  
 XX CC triglyceride content. The products can also be used to identify  
 XX CC antagonists and agonists of DAGAT activity. Such agonists and  
 XX CC antagonists are particularly useful in treating or ameliorating  
 XX CC diseases associated with DAGAT activity, including diseases associated  
 XX CC with altered cellular diacylglycerol concentration or PKC activity,  
 XX CC including cancer, diabetes, cardiopulmonary diseases e.g. heart failure,  
 XX CC atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,  
 XX CC metabolic disorders, obesity, diseases associated with abnormal lipid  
 XX CC metabolism, and diseases associated with abnormal fat absorption,  
 XX CC lipoprotein secretion and adipogenesis.  
 XX SQ Sequence 520 AA;  
 Query Match 100.0%; Score 2771; DB 21; Length 520;  
 Best Local Similarity 100.0%; Pred. No. 1.le-272;  
 Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MAILDSAGVTVTENGSGEFDLRLRRKRSRSDSSNGLLLSGSDNNSPDDVGAPADVR 60  
 DB 1 MAILDSAGVTVTENGSGEFDLRLRRKRSRSDSSNGLLLSGSDNNSPDDVGAPADVR 60  
 QY 61 DRIDSVNDAAQGTANLAGDNGGDNNGGCGGEGGNADFTYRPSVPAHRRARES 120  
 DB 61 DRIDSVNDAAQGTANLAGDNGGDNNGGCGGEGGNADFTYRPSVPAHRRARES 120  
 QY 121 PLSSDAIFKQSHAGLFNLCVVLVAVNSRLIENLMKYGLIRDTDFWSSRLSDWPLFM 180  
 DB 121 PLSSDAIFKQSHAGLFNLCVVLVAVNSRLIENLMKYGLIRDTDFWSSRLSDWPLFM 180  
 QY 181 CCISLSIFPLAAFTVEKLVLQKYLSEPVIFLHIIITMTVEVLYPVYTLRCDSAFLSGVT 240  
 DB 181 CCISLSIFPLAAFTVEKLVLQKYLSEPVIFLHIIITMTVEVLYPVYTLRCDSAFLSGVT 240  
 QY 241 LMLTCTVWLKLVSAHTSDIRSLANAADKANPEVSYVSLKSLAFVMAPTLCYOPSY 300  
 DB 241 LMLTCTVWLKLVSAHTSDIRSLANAADKANPEVSYVSLKSLAFVMAPTLCYOPSY 300  
 QY 301 PRSACIRKGGWVARQAKLVITFTGPMGTIEEQYINPIVRNSKHPKLGDLXIAIERVLKLSV 360  
 DB 301 PRSACIRKGGWVARQAKLVITFTGPMGTIEEQYINPIVRNSKHPKLGDLXIAIERVLKLSV 360

QY 361 PNLVYVLCMEFYCFPHLWNLITLAEELLCFCGDFREYKDWNAKSVGDIYRWNNMPVHKWVRH 420  
 DB 361 PNLVYVLCMEFYCFPHLWNLITLAEELLCFCGDFREYKDWNAKSVGDIYRWNNMPVHKWVRH 420  
 QY 421 IYFPCLSRKIPKTLAIIIAFLVSAVFHELCLIAVPCRLFKLWAFGLGIMFQVPLVITNYLQ 480  
 DB 421 IYFPCLSRKIPKTLAIIIAFLVSAVFHELCLIAVPCRLFKLWAFGLGIMFQVPLVITNYLQ 480  
 QY 481 ERFGSTVGNMIFWFIFCFIIFGQPMCVLLYYHDLNMRKGSMS 520  
 DB 481 ERFGSTVGNMIFWFIFCFIIFGQPMCVLLYYHDLNMRKGSMS 520

## RESULT 5

AAU00462  
 ID AAU00462 standard; Protein; 520 AA.

AC AAU00462;

DT 31-MAY-2001 (first entry)

XX Arabidopsis thaliana sterol acyltransferase ACAT.

KW Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;  
 KW acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;  
 KW nutritional supplement; dairy product; food product; salad dressing.

OS Arabidopsis thaliana.

PN W0200116308-A2.

XX 08-MAR-2001.

XX 30-AUG-2000; 2000WO-US23863.

XX 30-AUG-1999; 99US-0152493.

XX (MONS ) MONSANTO CO.

PI Lassner M, Van Eenennaam A;

DR WPI; 2001-169010/17.

DR N-PSDB; AAS01106.

PT New isolated nucleic acid encoding plant lecithin:cholesterol  
 PT acyltransferase-like or acyl (coenzyme A) CoA:cholesterol  
 PT acyltransferase-like polypeptides, for modifying the sterol content and  
 PT oil production of plants -

PS Example 3; Page 103-104; 127pp; English.

XX The present sequence represents Arabidopsis thaliana  
 CC acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel  
 CC polynucleotides encoding the plant sterol acyltransferases LCAT  
 CC (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341)  
 CC and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open  
 CC reading frame, LROI gene sequence (AAS01342), and a rat ACAT (AAS01105)  
 CC cDNA sequence are also described. The polynucleotides encoding LCAT  
 CC or ACAT are used to produce LCAT or ACAT polypeptides. They can also  
 CC be used in a recombinant construct to transform a host cell (preferably  
 CC of a plant) or a plant. The recombinant construct is used to increase or  
 CC decrease the sterol content of the host cell or plant. It can be used to  
 CC alter oil production of the cell or plant, preferably by increasing it.  
 CC The oil of the plant or the plant itself is used as a food product, or  
 CC as nutritional or dietary supplements, or in pharmaceutical compositions  
 CC for lowering cholesterol. The oil can be used in foods e.g. margarine,  
 CC butter, cooking oil, and dressings e.g. salad dressings, mayonnaise,  
 CC cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips,  
 CC baked goods, pastries, cookies, snack bars, confections, chocolates, and  
 CC beverages. The alteration in sterol content and/or composition can also  
 CC provide a plant with tolerance to stress and insect damage.

XX New DNA encoding diacylglycerol acyltransferase from *Arabidopsis*  
PT *thaliana* for transforming plants and regulating seed oil content, fatty  
PT acid synthesis and seed oil acyl composition in commercial and crop  
PT

Genetically engineering the biosynthetic pathways in plants involved in the accumulation of sterol compounds and tocopherol to produce compounds for lowering the level of low density lipoprotein cholesterol in blood serum

Disclosure: Page 58-59; 166pp; English.

The present sequence is that of Arabidopsis thaliana acyl CoA:cholesterol acyltransferase (ACAT), as deduced from a full-length ACAT DNA sequence (see AA88835). Sterol O-acyltransferases such as ACAT catalyze the formation of cholesterol esters from cholesterol and long chain fatty acids. Recombinant constructs of the invention are used to alter the biosynthesis and accumulation of sterols and tocopherols in transgenic plants. Seeds of such plants may contain elevated levels of sitosterol and/or its esters, and alpha-tocopherol, and reduced levels of campesterol and campestanol and their esters. The seeds may also contain the novel sterol brassicasterol. Oil obtained from the seeds can be used in food and pharmaceutical compositions to lower levels of low density lipoprotein cholesterol in blood serum. ACAT enzymes can be used in the present invention to produce elevated levels of phytosterol and/or phytosterol esters.

Sequence 520 AA;

```

Query Match      100.0%; Score 2771; DB 21; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.1e-272;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
DB 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
QY 61 DRIDSVDNDAQGTANLAGDNGGDNNGGCGGEGRGNADATFTYRPSVPAHRRARES 120
DB 61 DRIDSVDNDAQGTANLAGDNGGDNNGGCGGEGRGNADATFTYRPSVPAHRRARES 120
QY 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180
QY 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
DB 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
QY 241 LMLTCTIVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTLCYQPSY 300
DB 241 LMLTCTIVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTLCYQPSY 300
QY 301 PRSACIRKGWVARQFAKLVIFTFMGFTIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360
DB 301 PRSACIRKGWVARQFAKLVIFTFMGFTIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360
QY 361 PNLVYVLCMYCPFHLMNLAELLCFGDREFYKDWNNAKSVGDYWRWNMNPVHKWVRH 420
DB 361 PNLVYVLCMYCPFHLMNLAELLCFGDREFYKDWNNAKSVGDYWRWNMNPVHKWVRH 420
QY 421 IYFPLRSKIPKTLAIIIAELVSADFELCTIAPCRFLKFAFLGIMFQVPLVITNLYQ 480
DB 421 IYFPLRSKIPKTLAIIIAELVSADFELCTIAPCRFLKFAFLGIMFQVPLVITNLYQ 480
QY 481 ERFGSTVGNMIFWFIIFCIGQPMCVLLYYHDLNLRKGSMS 520
DB 481 ERFGSTVGNMIFWFIIFCIGQPMCVLLYYHDLNLRKGSMS 520

```

RESULT 2

AA94524

ID AA94524 standard; Protein; 520 AA.

XX

AC AA94524;

XX

DT 06-DEC-2000 (first entry)

XX DE A. thaliana diacylglycerol acyltransferase protein variant #1.

XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX KW triacylglycerol; herbicide; EC2.3.1.20.

XX OS Arabidopsis thaliana.

XX PN WO200032756-A2.

XX PD 08-JUN-2000.

XX PF 01-DEC-1999; 99WO-US28354.

XX PR 02-DEC-1998; 98US-0110602.

XX PR 31-MAR-1999; 99US-0127111.

XX PA (DUPO) DU PONT DE NEMOURS & CO E. I.

XX PI Cahoon EB, Kinney AJ, Cahoon RE;

XX DR WPI; 2000-412308/35.

Polynucleotides encoding diacylglycerol acyltransferase, useful for synthesis of triacylglycerols and increasing the level of oils in plant seeds

Example 4; Fig 1; 62pp; English.

In the present invention, cDNA libraries from Arabidopsis, corn, rice, soybean and wheat were screened for sequences with homology to a putative acyl CoA cholesterol acyltransferase related gene from Arabidopsis thaliana and diacylglycerol acyltransferases from Homo sapiens and Mus musculus. The cDNA clones identified from this process were used to form complete diacylglycerol acyltransferase cDNA sequences. The present sequence is Arabidopsis thaliana diacylglycerol acyltransferase protein. This sequence was used for homology comparison with the novel diacylglycerol acyltransferases. Diacylglycerol acyltransferases are involved in the synthesis of triacylglycerols. Alteration of the expression of the diacylglycerol acyltransferase DNA can be useful for increasing the level of oils in plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful as herbicides.

Sequence 520 AA;

```

Query Match      100.0%; Score 2771; DB 21; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.1e-272;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
DB 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
QY 61 DRIDSVDNDAQGTANLAGDNGGDNNGGCGGEGRGNADATFTYRPSVPAHRRARES 120
DB 61 DRIDSVDNDAQGTANLAGDNGGDNNGGCGGEGRGNADATFTYRPSVPAHRRARES 120
QY 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180
QY 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
DB 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
QY 241 LMLTCTIVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTLCYQPSY 300
DB 241 LMLTCTIVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTLCYQPSY 300
QY 301 PRSACIRKGWVARQFAKLVIFTFMGFTIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360
DB 301 PRSACIRKGWVARQFAKLVIFTFMGFTIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:48:34 ; Search time 71 Seconds

(without alignments)  
1162.504 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTWGGGEF.....QPMCVLLYYHDLNRKSGMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_19Jun03.\*

```

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	21	Arabidopsis acyl C
2	2771	100.0	520	21	A. thaliana diacyl
3	2771	100.0	520	21	A. thaliana diacyl
4	2771	100.0	520	21	Acyl-CoA:cholester
5	2771	100.0	520	22	Arabidopsis thalia
6	2753	99.4	520	21	Arabidopsis diacyl
7	2747.5	99.2	547	21	A. thaliana AS11 m
8	1821	65.7	504	21	Soybean diacylglyc
9	1636.5	59.1	500	21	Rice diacylglycerol

10	1596	57.6	508	21	AA194522	Wheat diacylglycer
11	1211	43.7	361	21	AA194513	Corn diacylglycerol
12	1046.5	37.8	327	21	AA194515	Corn diacylglycerol
13	1030.5	37.2	285	21	AA194516	Corn diacylglycerol
14	791.5	28.6	488	21	AA15200	Human ACAT Related
15	773.5	27.9	500	21	AA13742	Rat acyl CoA:chole
16	773.5	27.9	500	21	AA154139	Acyl-CoA:cholester
17	770.5	27.8	497	21	AA194523	Mouse diacylglycer
18	770.5	27.8	498	21	AA144562	Mouse diacylglycer
19	760.5	27.4	489	23	AAE24973	Bovine DGAT1 prote
20	758.5	27.4	467	23	AAE24974	Bovine diacylglyce
21	758.5	27.4	489	24	ABP96046	Bovine DGAT protei
22	756.5	27.3	489	23	AAE24979	Bovine diacylglyce
23	754.5	27.2	489	24	ABP96045	Bovine DGAT protei
24	738.5	26.7	407	19	AAW43406	Human acylcoenzyme
25	724.5	26.1	496	21	AA19741	Caenorhabditis ele
26	724.5	26.1	496	21	AA154140	Acyl-CoA:cholester
27	696.5	25.1	386	21	AA144561	Human diacylglycer
28	612	22.1	155	21	AA194514	Corn diacylglycerol
29	573.5	20.7	236	22	ABB10191	Human cDNA SEQ ID
30	573.5	20.7	236	23	ABP66778	Human polypeptide
31	526.5	19.0	219	19	AAW43412	Human acylcoenzyme
32	397	14.3	550	19	AAW43409	Human acylcoenzyme
33	396	14.3	550	15	AA153079	Acetyl coenzyme A:
34	396	14.3	550	19	AAW38416	Human acyl-coenzym
35	378.5	13.7	642	19	AAW43411	Yeast acylcoenzyme
36	375.5	13.6	642	19	AAW38418	Yeast acyl-coenzym
37	375	13.5	525	21	AA167952	Mouse acyl CoA:cho
38	374	13.5	559	22	ABP63696	Drosophila melanog
39	357.5	12.9	522	21	AA121640	Human ACAT Related
40	351	12.7	502	21	AA167953	Human acyl CoA:cho
41	340.5	12.3	165	22	AAU17635	Novel signal trans
42	307.5	11.1	80	21	AA194517	Rice diacylglycerol
43	301	10.9	610	19	AAW43410	Yeast acylcoenzyme
44	301	10.9	610	22	AA197263	Yeast ARE1 amino a
45	299	10.8	610	19	AAW38417	Yeast acyl-coenzym

#### ALIGNMENTS

```

RESULT 1
AA191740
ID  AA191740 standard; Protein; 520 AA.
XX
AC  AA191740;
XX
XX
DT  19-FEB-2001 (first entry)
XX
DE  Arabidopsis acyl CoA:cholesterol acyltransferase.
XX
XX  Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
KW  sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
KW  hypolipemic.
XX
XX  Arabidopsis thaliana.
XX
XX  WO2000061771-A2.
XX
XX  19-OCT-2000.
XX
XX  12-APR-2000; 2000WO-US09696.
XX
XX  12-APR-1999; 99US-0128995.
XX
XX  (MONS ) MONSANTO CO.
XX
XX  Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;
XX  Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;
XX  WPI: 2000-665136/64.
XX  N-PSDB; AAA88835.

```





ADDRESS: LAHIVE & COCKFIELD  
STREET: 28 STATE STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/121  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/509  
FILING DATE: 07/31/95  
ATTORNEY/AGENT INFORMATION:  
NAME: LAMPORT HAMMITTE, ANN.  
REGISTRATION NUMBER: 34,858  
REFERENCE/DOCKET NUMBER: DGI-  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-2700  
TELEFAX: (617) 743-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 550 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-121-396-4

Query Match 14.3%; Score 396; DB 2; Length 550;

17	QY	GGBEVDLRLRRKSRSDSNGL-----LLSGSDNNSPSDVVGAPADVRDRIDSVNDD	70
70	Db	GSFHDDPVTWLIERSASLDNGGCALTTFVSLEGEKNHRAKDLRAPPE-----	117
71	QY	AQGTANLAGNNGGDDNNGCGGCGEGRNADATFTTTPSPVPAHRRARESPLSSDAIPKQ	130
118	Db	-QCKIFT-----ARRSLIDELLEWDHIRT	141
131	QY	SHAGFLNLCVVVLIIV-----NSRLIIE-NLMKYG-LIIRDFWESSRSLRDPLEFMC	182
142	Db	YHMFIALLIIFILSTLVVDYIDEGRLVLEFSLSYAFGKPTTVVW-----TWIMF----	192
183	QY	ISLSIPPLAAFTVEKVLQKY-----ISEPVI-----FLHIITWTEVLY-PVYVTLR	230
193	Db	-----LSFSPYFLFQHWRTGYSKSSHPILRSFLHGFLEMLFGVLGFGTYVVL-	244
231	QY	CDSAFISGVITMLLCTIVLWKLVSYAHT-----SYDIRSLANAADKAN-----PEVSYYVSLK	283
245	Db	---AYTLPPASRFTIIEQIRFVMKAHSFVRENVPVLSNAKESSTVPIPTVNOYL----	298
284	QY	SLAYEVAPLTCQPSYPRACIRKGHWARQAKLVITFGMGFIIEQYINPIVRNSK-H	342
299	Db	---YLFAPLTIYRDSYPRNPTVWGVAMKFQVFCGFYVYIFERLCAPLERFNKQE	355
343	QY	PLKGDLLYATERVKLSV-----PNLVWLCMFYCFPHLNLIAELLFCGDRFYKDW	397
356	Db	-----RVILVLCVFNLSILPGVILILFLTFFAFLHCLNFAEMLRFGDRMFYKDW	408
398	QY	NAKSVGDYWRMNNPWHKWMVRHIY---FPCLRSKIPKTLAIIIAFLVSAPVHELCAVPC	455
409	Db	NSTSYSNYRTWNVVHMDLVYYAYKDFLWFFSKRKFSAAMLAVFAVSAVVHEYALAV-C	467
456	QY	R-----LFKLWFLGIMFOVPLVFTINYLOERPGSTVGNMIFW-----FTFCIFGQ	501
468	Db	LSFFYPVLFLVFMFFGMAFN-----FIVNDSRKK---PIWVLMWTSFLGLNGVLLCYSQ	520

```

RESULT 13
PCT-US93-09704A-4
; Sequence 4, Application PC/TUS9309704A
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF INVENTIONS: 9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09704A
; FILING DATE: October 12, 1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. SER. NO. 959,950
; FILING DATE: October 14, 1992
; APPLICATION NUMBER: U.S. SER. NO. 121,057
; FILING DATE: September 10, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMPORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 550 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-09704A-4

```

```

Query Match      14.3%; Score 396; DB 5; Length 550;
Best Local Similarity 26.7%; Pred. No. 1.5e-30;
Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

QY 17 GGEFVDLRLRRKRSRSDSNGL-----LLSGSDNNSPSDVGADPADVRDRIDSVVDD 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 70 GSHFDDFVTLIEKSASLONGGCGALTFTFSLEGEKNHRAKDLRAPPE----- 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 71 AQTANLAGNNGGGDNGGGRGGGGRGNADATFTYRPSVPAHRRARESPSSDAIFKQ 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 118 -QKRII-----ARRSLDELLEVDHIRT 141
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 131 SHAGLENLGVVLIAY-----NSRLIIE-NLMKGW-LIRTDWFSRSLRDWPUMCC 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 YHMFIALILFILTSLVDYIDEGRLVLEFSLSYAFGEPTVW-----TWIMF--- 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 183 ISLSIEPLAAFTVEKLVLOKY-----ISEPVI-----FLHIITMTEVY-PTYVTLR 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 193 -----LSTFSVPYELFQHWRTGYSSKSHPLIRSLFHGHFLMFIQIGVLGFGPTVVVL- 244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 231 CDSAFISGVYTLMLTICIVWLKLVSYAHT---SYDIRSLANRADKAN---PEVSYTVSK 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 ---AYTLPPASRIIIFEQIRFMKASHVRENVPRLVNSAKESSVTPIPTVNOYL--- 298
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 284 SLAYFVWAPLTCYOPSPRSACIRKGWAROAKLVITGFMGFIIEQYINPTVRNKS-H 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 ---YFLFAPTLIYRDSYPRNPVWGYVAMKFAQVFGCGFYVYIFERLCAPLFRNIKOE 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 343 PLKGDLLYAIERYKLVS-----PNLYVWLCMFYCFPHLWNLAIELLFCGFREFYKDW 397
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 356 PFSA-----RVVLVCFVNSILPGVLIILFLFAFLHCLWLNFAEMLRFGDMFYKDW 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 398 NAKSGVDYWRMNMPPVHKWVRHIY--FPLCRSKIPKTLAIATIAFLVSAVFHELCTAVPC 455
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 409 NSTSYGNYRTNVVVDHLYYYAYKDFLWFFSKRFSKSAAMLAVFAVSAAVHEYLAV-C 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-2700  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 550 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-121-057-4

Query Match 14.3%; Score 396; DB 1: Length 550;  
Best Local Similarity 26.7%; Pred. No. 1.5e-30;  
Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;  
QY 17 GGEFVLDLRLRRKSRSDSNGL-----LLSGDNNSPDDVGPADYRDRIDSVVND 70  
DB 70 GSHFDDFVNLIKESASLNGGCAITTFVLEGEKNNHRAKDLRAPPE----- 117  
QY 71 AQTANLAGDNGGGGNGGGRGGGGRGNADATFTYRPSVPAHRRARSPSSDAIFKQ 130  
DB 118 -OGKIFI-----ARRSLDELLEVDHRTI 141  
QY 131 SHAGLENLGVVLIIV-----NSRLIE-NLMKYGW-LIRTDWFSSRLRDPFMCC 182  
DB 142 YHMFIALILFILTSTLVVDYIDEGRLVLEFSLSYAFGKPTVVW-----TWIMF 192  
QY 183 ISLSIFPLAAFTVEKVLQKY-----ISEPVVI-----FLHIIITMTEVLY-PVYVTLR 230  
DB 193 -----LSTFSPVYFLFQHWRTGYSKSHPLRSFLHGFLEFQIGVLGFGPTVVL- 244  
QY 231 CDSAFSLGVTMLLTICVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283  
DB 245 ---AYLTPASRFIIIFQIRFVMAHSEVRENVPRLNSAKEKSTVPIPTVQYL--- 298  
QY 284 SLAYFVAPTLCTQPSYPSRACIRKGWARQAKLVITGFMGFIIEQYINPIVRNSK-H 342  
DB 299 ---YLFAPTLIYRDSYPRNPVTVWGYVAMKFAQVFCGCFYVYIIFERLCAPLFRNIKE 355  
QY 343 PLKGDLLYAIERVLKLSV-----PNLYVWLCMFYCFHHLNLTLAELLCGDREFYKDW 397  
DB 356 PFSA-----RVLVLCVFNILPGVLLFLTFPAFLHCLWNAFAEMLRFGDRMFYKDW 408  
QY 398 NAKSVGDYWRMNMVHKWVRHIY--FPCLRKIPKTLAIIITAFVLSAVFHELCTAVPC 455  
DB 409 NSTSYNYRTNMVVDWLYIYAYKDFLWFFSKRFSKSAAMLAFAVAVVHEYALAV-C 467  
QY 456 R-----LFKLWAFGLIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIGQ 501  
DB 468 LSEFFYPVLVLFMFFGMAFN---FIVNDSRKK---PIWNLMTSLFLGNGVLLCFYSQ 520

RESULT 11  
US-08-509-187D-4  
Sequence 4, Application US/08509187D  
Patent No. 5834283  
GENERAL INFORMATION:  
APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.  
TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/509,187D  
FILING DATE: 31-JUL-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Lampert Hammitte, Ann  
REGISTRATION NUMBER: 34,858  
REFERENCE/DOCKET NUMBER: DCI-033cpdv  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 742-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 550 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-509-187D-4

Query Match 14.3%; Score 396; DB 2: Length 550;  
Best Local Similarity 26.7%; Pred. No. 1.5e-30;  
Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;  
QY 17 GGEFVLDLRLRRKSRSDSNGL-----LLSGDNNSPDDVGPADYRDRIDSVVND 70  
DB 70 GSHFDDFVNLIKESASLNGGCAITTFVLEGEKNNHRAKDLRAPPE----- 117  
QY 71 AQTANLAGDNGGGGNGGGRGGGGRGNADATFTYRPSVPAHRRARSPSSDAIFKQ 130  
DB 118 -OGKIFI-----ARRSLDELLEVDHRTI 141  
QY 131 SHAGLENLGVVLIIV-----NSRLIE-NLMKYGW-LIRTDWFSSRLRDPFMCC 182  
DB 142 YHMFIALILFILTSTLVVDYIDEGRLVLEFSLSYAFGKPTVVW-----TWIMF 192  
QY 183 ISLSIFPLAAFTVEKVLQKY-----ISEPVVI-----FLHIIITMTEVLY-PVYVTLR 230  
DB 193 -----LSTFSPVYFLFQHWRTGYSKSHPLRSFLHGFLEFQIGVLGFGPTVVL- 244  
QY 231 CDSAFSLGVTMLLTICVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283  
DB 245 ---AYLTPASRFIIIFQIRFVMAHSEVRENVPRLNSAKEKSTVPIPTVQYL--- 298  
QY 284 SLAYFVAPTLCTQPSYPSRACIRKGWARQAKLVITGFMGFIIEQYINPIVRNSK-H 342  
DB 299 ---YLFAPTLIYRDSYPRNPVTVWGYVAMKFAQVFCGCFYVYIIFERLCAPLFRNIKE 355  
QY 343 PLKGDLLYAIERVLKLSV-----PNLYVWLCMFYCFHHLNLTLAELLCGDREFYKDW 397  
DB 356 PFSA-----RVLVLCVFNILPGVLLFLTFPAFLHCLWNAFAEMLRFGDRMFYKDW 408  
QY 398 NAKSVGDYWRMNMVHKWVRHIY--FPCLRKIPKTLAIIITAFVLSAVFHELCTAVPC 455  
DB 409 NSTSYNYRTNMVVDWLYIYAYKDFLWFFSKRFSKSAAMLAFAVAVVHEYALAV-C 467  
QY 456 R-----LFKLWAFGLIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIGQ 501  
DB 468 LSEFFYPVLVLFMFFGMAFN---FIVNDSRKK---PIWNLMTSLFLGNGVLLCFYSQ 520

RESULT 12  
US-09-121-396-4  
Sequence 4, Application US/09121396  
Patent No. 5968749  
GENERAL INFORMATION:  
APPLICANT: CHANG, TA-YUAN  
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL  
TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:

ATTORNEY/AGENT INFORMATION:  
NAME: Field, Bret E  
REGISTRATION NUMBER: 37,620  
REFERENCE/DOCKET NUMBER: 6510-105p  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650 327 3400  
TELEFAX: 650 327 3231  
TELEX:  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-103-754A-4

Query Match 25.1%; Score 696.5; DB 4; Length 386;

Best Local Similarity 40.9%; Pred. No. 1.7e-60;

Matches 157; Conservative 68; Mismatches 126; Indels 33; Gaps 12;

QY 150 LIENLMRYGLIRDFWSSRLRD---WPLFMCCISLIFPLAAFTVEKLVLOKYE 206

Db 1 LELENLIYGLV-DPIQVSLFLKDPYSPAPCLVIAANFVAQVERKLVGALTE 59

QY 207 PVVIFLIIITWTEVLYPVVYTLRCDSAFLSGVTMLLT-CIVWLKLVSYAHTSYDIRS- 264

Db 60 QAGLLHLVANLATILCPFAAVVLLVESITPVGSLALMAHTILFLKLSYR----DVNSW 115

QY 265 -----LANAADRANP-----EVSY--YVSLKSLAYFWVAPTICQPSYPSACIRKGW 310

Db 116 CRRARAKASAGKASSVAAPHTVSYDNLTYRDLTYFLFAPTLCYELNFRSPRIKRF 175

QY 311 VARQAKLVITGFMGFIIEQYINPIVRNSKHPKLGDLIYA--IBRVKLSVPNLYVWLC 368

Db 176 LLRLLEMLFTQLQVGLIQWVPTIONSMPKPK-DMYSRIERLLKLVNHLIWI 234

QY 369 MYFCFFHMLNLAELLCFGDREFFYKQWNNKSVGDYWRMNMVHKWVRHIYPPCLRS 428

Db 235 FFYWLFSCLNAVALMQDREFFYRDMWNSVTFYFQNNIPVHKWCIRHFYKPMRLR 294

QY 429 KIPKTLAIIIAFLYSAPVHELCLIAVPCRLFLKLAFLGIMFQVPLV-FITNVLQERFGSTV 487

Db 295 GSSKWMATGTVLASAFFHEHLYSVPLRFLWAFMGMAQIPLAWFVGRFFQNGY---- 350

QY 488 GNMIFWFICFGOPMVCVLLYYHD 511

Db 351 GNAAVW-LSLIIGQPIAVLMTVHD 373

## RESULT 9

US-09-326-203A-23

; Sequence 23, Application US/09326203A

; Patent No. 644876

; GENERAL INFORMATION:

; APPLICANT: Lassner, Mike

; APPLICANT: Ruzinsky, Diane

; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic

; FILE REFERENCE: 17045/00/WO

; CURRENT APPLICATION NUMBER: US/09/326,203A

; CURRENT FILING DATE: 1999-06-04

; PRIOR APPLICATION NUMBER: 60/088,143

; PRIOR FILING DATE: 1998-06-05

; PRIOR APPLICATION NUMBER: 60/108,389

; PRIOR FILING DATE: 1998-11-12

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 23

; LENGTH: 409

; TYPE: PRT

; ORGANISM: murine

; PUBLICATION INFORMATION:

; REFERENCE/DOCKET NUMBER: DCI-033CP

; ATTORNEY/AGENT INFORMATION:

; NAME: LAMPORT HAMMITTE, ANN.

; REGISTRATION NUMBER: 34,858

; REFERENCE/DOCKET NUMBER: DCI-033CP

; JOURNAL: J. Biol. Chem.

; VOLUME: 270

; PAGES: 26192-26201

; DATE: 1995

US-09-326-203A-23

## Query Match

Best Local Similarity 29.4%; Pred. No. 2e-32;

Matches 129; Conservative 73; Mismatches 139; Indels 98; Gaps 23;

QY 118 RESPLSSDAIFKQSHAG-----LFNLCVVVLAVN-SRLIIE-NLMKYGW-LIR 163

Db 1 ROSLL--DELFEVDHIRTYYHMFIALLLILVLTSTVDYIDEGRLVLEFNLLAYAGRFP 58

QY 164 TDFWFSRSLRDWPLFMCCISLIFPLAAFTVEKLVLOKYE-----SEPVIIFL-HII 216

Db 59 TVIW-----TWAMFLSTLSIPY-----LFQPAHGYSKSHPLIYSLVHGLL 102

QY 217 TWTEVL-----YPVYVTLRCDSAFLSGVTMLLT-CIVWLKLVSYAHT--SYDIRSLANNA 269

Db 103 FLVFGVLGVPVTVVWL---AYLPPASRFLILEQIRLIMKAHSFVRENIPLVNA 158

QY 270 DKAN-----PEVSYVYVSLKSLAYFWVAPTICQPSYPSACIRKGWVARQPAKLVIFTCF 324

Db 159 KEKSKDPLPTVNOYL-----YFLFAPTLYRDYPRTPVTRWGYVAMQFLQVFGCLFY 212

QY 325 MGFIIEQYINPIVRNSK-HPLKGDLLYAIAERVLKLSV-----PNLYVWLCHMFCFHLWL 378

Db 213 VYIYFELCAPLFRNIKQEPFSA-----RVVLVLCVNSILPGVLIILFSEFAFLHCWL 265

QY 379 NILAELLCFGDREFFYKQWNNKSVGDYWRMNMVHKWVRHIYPPCL--RSKIPKTLAI 436

Db 266 NAFADMLRFGDMFYKDMWNSTSYNYRTWNVVVDWLYVYVYKDLLWFFSKRFSKSAAM 325

QY 437 IIAFLVSAPVHELCLIAVPCRLFLKLAFLGIMFQVPLV-FITNVLQERFGSTVGNM 490

Db 326 LAVFALSAAVHAYALAI-CLSYTFYVPLVFLVFMFGMAFN----FIVNDSRRK---PIWNI 377

QY 491 IFW-----FIFCIFGQ 501

Db 378 MWASLFLGYGLILCFYSQ 396

## RESULT 10

US-08-121-057-4

; Sequence 4, Application US/08121057

; Patent No. 5484727

; GENERAL INFORMATION:

; APPLICANT: CHANG, TA-YUAN

; APPLICANT: CHANG, CATHERINE C. Y.

; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL

; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD

; STREET: 60 STATE STREET, SUITE 510

; CITY: BOSTON

; STATE: MA

; COUNTRY: USA

; ZIP: 02109

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/121.057

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: LAMPORT HAMMITTE, ANN.

; REGISTRATION NUMBER: 34,858

; REFERENCE/DOCKET NUMBER: DCI-033CP



Query Match 28.6%; Score 791.5; DB 3; Length 488;  
Best Local Similarity 38.0%; Pred. No. 9.7e-70;  
Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;

QY 25 RLRRKSRSSSSNGLLSGSDNNSPSDDVGAPADVRIDSVNDDAQ--TANLAGDNG 83  
DB 8 RRRRTGGRPSHG-----GGPAAAE-----VRDAAAGPDVGAAGDAPA 48  
QY 84 GGDNNGGGRGGEGGRGNADATFTYRSPVAPHRRAAREPLSSDAIFKQSHAGLNLGVVL 143  
DB 49 PAPNKGDAGVGSCHWELR-----HRLQDSLFSDSGF--SNYRILNKCVMVL 96  
QY 144 IAVNSRLIENIMKYGMIRDFWFFSSRLRD---WPLFMCCISLSIFPLAAFTVEKLV 200  
DB 97 ILSNARLFLENLIKYGILV-DPIQWLSLFKDPHSPAPCLVIAANVFAVAAFOVEKRLA 155  
QY 201 QKYISEPVFHLIIITWEVLYPVYVTLRCDSAFLSGVTLMLIT-CIVWLKLVSYAHTS 259  
DB 156 VGLTEAQEGLLHVNANLATICFPAAVVLVESITPVGSLALMAHTILFLKFSYR--- 212  
QY 260 YDIRS-----LANAADKANP-----EVSY--YVSLKSLAYFMVAPTLCYQSPBSA 304  
DB 213 -DVSNCRRARAKAASAGKASSAAAPHTSYDNLTYRDIYLFELPTCYELNFPSP 271  
QY 305 CIRKGVAROFKLVITFGMFTIEQVINIVRNSKHLPGKDLLYA--IERVLKLSVPN 362  
DB 272 RIRKRELLRLEMLFTQLQVGLIQMWPTIONSMPFK-DMDYSRIERLLKLVYPN 330  
QY 363 LYVWCMFYCFHLMNLIELLCFGDREFFYKDWNAKSGVDYWRMMNMPVHKWVRHIY 422  
DB 331 HLIWLIFFYWLFSCLNAVALMQLQGDREFFYRDMWNSSEVTFYQWNNIPVHKWCIRHFY 390  
QY 423 FPLRSRIPKTLAIIIFLVSAPHELCIAPCRFLKWLAFGLIMFQVPLV-FITNYLQE 481  
DB 391 KPLRRGSSKWMARTGVFLASAFFHEYLVSPLRMFLWAFMGMAQIPLAWFVGRFFQG 450  
QY 482 RFGSTVGNMIFWFICIFGPMCVLLYYHD 511  
DB 451 NY----GNAAVW-LSLIIGQPIALMYVHD 475

RESULT 5  
US-09-326-203A-17  
; Sequence 17, Application US/09326203A  
; Patent No. 644876  
; GENERAL INFORMATION:  
; APPLICANT: Lassar, Mike  
; APPLICANT: Ruezinsky, Diane  
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic  
; TITLE OF INVENTION: Acid Sequences  
; FILE REFERENCE: 17045/00/WO  
; CURRENT APPLICATION NUMBER: US/09/326, 203A  
; CURRENT FILING DATE: 1999-06-04  
; PRIOR APPLICATION NUMBER: 60/088,143  
; PRIOR FILING DATE: 1998-06-05  
; PRIOR APPLICATION NUMBER: 60/108,389  
; PRIOR FILING DATE: 1998-11-12  
; NUMBER OF SEQ ID NOS: 46  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Rattus sp.  
US-09-326-203A-17

QY	73	GTANLADGNNCGDNGCGGGEGGRGNADATFTYRPSVPAHRRARESPSLSDAIFKQSH
Db	68	TSV-----GDGHWEIR-----CH-RLQDSLFSDSGSF-SNY 96
QY	133	AGLFNLCVVLIAYNSRLIIENLMKYGLWLRDTDFWFSSRSIRD--WPLFMCCISLSIIPP 189
Db	97	RGIUNWCVMILLSNARLFLENLIKYGILV-DPIQVVSFLKDPYSWPAPCLIIASNIET 155
QY	190	LAAFTVEKLVLQKYZISEPVFIHLHIITMTVELYPVYVTLRCDSAFLSGVTMLLLT-CIV 248
Db	156	VATEOIEKRLSVGALTEQMGLLHVVNLATIICPPAAVALLVESITTPVCSLPALASYII 215
QY	249	WLKLVSVAHTSY-----DIRSLANAADK-----ANPEVSY--YYSLKSLAFMWAPTLC 295
Db	216	FLKFLSVDNLWNCQRORRVAKAVSAGRKKVGAAQAQNTSYSPDNETYRDLYYFIFAPTLC 275
QY	296	YOPSYPRXSACIRKGWAROFKLVFTCFMGFPIEBOYNPIVRNSKHPLKLGDLIVA--TE 353
Db	276	YELNFPSPRIKRKFLLRVLLEMFFTOQLQGLIQOMMVPTIQNSMKPK-KDMFYRIIE 334
QY	354	RVLKSLPNLYVWLCMFYCFPHLMLNILAEELCFGDFEFYKDWNAKSVDGYWRWMNMPV 413
Db	335	RLKLAVPNHLWLIIFFVWLFPHSCINAVAELLQFGDREYRDWNAAESVITYPQNKNIPV 394
QY	414	HKMVRHYIYPCRLSKPKTKTLAIIAFLVSADFHELCTAVPCRLEFKLWAFIMFOVPVLY 473
Db	395	HKMCIRHFYKPMRLGSKNMKMAQTGVFLASAFFHEYLVSILPMFRLWAFWTAMAOVPLA 454
QY	474	FITNVLQERFGSTGMNFWFIFCFIGOPMCVLLYYHD 511
Db	455	WIYNRF---FQNGYGNAAVW-VTLIGQPVAVLMYVHD 488
 RESULT 6 US-09-103-754A-5 ; Sequence 5, Application US/09103754A ; Patent No. 6344548 ; GENERAL INFORMATION: ; APPLICANT: Farese, Robert ; APPLICANT: Cases, Sylvaine ; APPLICANT: Smith, Steven ; APPLICANT: Erickson, Sandra ; TITLE OF INVENTION: Diacylglycerol O-acyltran ; TITLE OF INVENTION: sferase ; NUMBER OF SEQUENCES: 6 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: Bozicevic & Reed ; STREET: 285 Hamilton Avenue, Suite 200 ; CITY: Palo Alto ; STATE: CA ; COUNTRY: USA ; ZIP: 94301 ; COMPUTER READABLE FORM: ; MEDIUM TYPE: Diskette ; COMPUTER: IBM Compatible ; OPERATING SYSTEM: DOS ; SOFTWARE: FastSeq for Windows Version 2.0 ; CURRENT APPLICATION DATA: ; APPLICATION NUMBER: US/09/103,754A ; FILING DATE: ; CLASSIFICATION: ; PRIOR APPLICATION DATA: ; APPLICATION NUMBER: ; FILING DATE: ; ATTORNEY/AGENT INFORMATION: ; NAME: Field, Bret E ; REGISTRATION NUMBER: 37,620 ; REFERENCE/DOCKET NUMBER: 6510-105p ; TELECOMMUNICATION INFORMATION: ; TELEPHONE: 650 327 3400 ; TELEFAX: 650 327 3231 ; TELEX:		

```
Db 301 PRSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIAERVLSV 360
QY 361 PNLYVWLCMFYCFHFWLNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVRH 420
Db 361 PNLYVWLCMFYCFHFWLNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVRH 420
QY 421 IYFPCLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLGIMFQVPLVFTNYLQ 480
Db 421 IYFPCLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLGIMFQVPLVFTNYLQ 480
QY 481 EREGSTVGNMIFWIFCFIFGQPMCVLLYYHDLNMRKGSMS 520
Db 481 EREGSTVGNMIFWIFCFIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 2
US-09-593-359-4
; Sequence 4, Application US/09593359
; Patent No. 6552250
; GENERAL INFORMATION:
; APPLICANT: Larocche, Andre J.
; APPLICANT: Nykiforuk, Cory L.
; APPLICANT: Weselake, Randall J.
; TITLE OF INVENTION: Diacylglycerol O-acyltransferase
; FILE REFERENCE: 24015050
; CURRENT APPLICATION NUMBER: US/09/593,359
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Brassica napus
; OTHER INFORMATION: DGAT1
US-09-593-359-4

Query Match 83.3%; Score 2309.5; DB 4; Length 503;
Best Local Similarity 84.88; Pred. No. 8.2e-220;
Matches 442; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 1 MAILDSAGVTT-VTENGGBFVDLRLRRKRSRSDSNGLLSGDNNSPSDVDGAPADY 59
Db 1 MAILDSGGVAVPTENG---VADLDRLHRRKSSDSNGLL---SDTSPSDVGAAGAAE 53
QY 60 RDRDSDVNDADGATNACDNGGDNNGGGGGGCGGNADATFYRPSVFAHRRARE 119
Db 54 RDRVDSAAEEAGGTANLA-----GGDAETRESAGG-----DVRFTYRPSVFAHRRTRE 102
QY 120 SPLSSDAIFKQSHAGFLNLCVVVLIIVNSRLIENLMKYGLWIRTDFWSSRSRLRDWPLF 179
Db 103 SPLSSDAIFKQSHAGFLNLCVVVLIIVNSRLIENLMKYGLWIRTDFWSSRSRLRDWPLF 162
QY 180 MCCISLSIFPLAFTVEKLVQKYESEPVVIFLHIIITMTTEVLYPVVTLRCDSAFSLGV 239
Db 163 MCCLSLSVFPLAFTVEKLVQKYESEPVVIFLHIIITMTTEVLYPVVTLRCDSAFSLGV 222
QY 240 TMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSKLSLAYFNVAPVTLCYQPS 299
Db 223 TMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSKLSLAYFNVAPVTLCYQPS 282
QY 300 YPRSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIAERVLSKLS 359
Db 283 YPRSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIAERVLSKLS 342
QY 360 VPVLYVWLCMFYCFHFWLNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 419
Db 343 VPVLYVWLCMFYCFHFWLNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 402
QY 420 HIYFPCRLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLGIMFQVPLVFTNYL 479
Db 403 HIYFPCRLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLGIMFQVPLVFTNYL 462
QY 480 QERFGSTVGNMIFWIFCFIFGQPMCVLLYYHDLNMRKGSMS 520
```

```
Db 463 QERFGSMVGNMIFGSASCIFGQPMCVLLYYHDLNMRKGSMS 503

RESULT 3
US-09-593-359-2
; Sequence 2, Application US/09593359
; Patent No. 6552250
; GENERAL INFORMATION:
; APPLICANT: Larocche, Andre J.
; APPLICANT: Nykiforuk, Cory L.
; APPLICANT: Weselake, Randall J.
; TITLE OF INVENTION: Diacylglycerol O-acyltransferase
; FILE REFERENCE: 24015050
; CURRENT APPLICATION NUMBER: US/09/593,359
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Brassica napus
; OTHER INFORMATION: DGAT2
US-09-593-359-2

Query Match 62.08; Score 1717; DB 4; Length 341;
Best Local Similarity 93.08; Pred. No. 1.8e-161;
Matches 317; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 180 MCCISLSIFPLAFTVEKLVQKYESEPVVIFLHIIITMTTEVLYPVVTLRCDSAFSLGV 239
Db 1 MCCLSLSVFPLAFTVEKLVQKYESEPVVIFLHIIITMTTEVLYPVVTLRCDSAFSLGV 60
QY 240 TMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSKLSLAYFNVAPVTLCYQPS 299
Db 61 TMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSKLSLAYFNVAPVTLCYQPS 120
QY 300 YPRSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIAERVLSKLS 359
Db 121 YPRSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIAERVLSKLS 180
QY 360 VPVLYVWLCMFYCFHFWLNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 419
Db 181 VPVLYVWLCMFYCFHFWLNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 240
QY 420 HIYFPCRLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLGIMFQVPLVFTNYL 479
Db 241 HIYFPCRLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLGIMFQVPLVFTNYL 300
QY 480 QERFGSTVGNMIFWIFCFIFGQPMCVLLYYHDLNMRKGSMS 520
Db 301 QERFGSMVGNMIFGSASCIFGQPMCVLLYYHDLNMRKGSMS 341

RESULT 4
US-09-165-042-1
; Sequence 1, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
; APPLICANT: Sturley, Stephen L.
; APPLICANT: Oelkers, Peter
; TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
; FILE REFERENCE: 0575/56331
; CURRENT APPLICATION NUMBER: US/09/165,042
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Yeast
US-09-165-042-1
```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 04:55:04 ; Search time 43 Seconds  
(without alignments)  
511.666 Million cell updates/sec

Title: US-09-623-514a-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTTTENGGEF.....OPMCVLLYHDLNKRKGSMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2.6/ptodata/1/1aa/5A\_COMB.pep.\*
- 2: /cgn2.6/ptodata/1/1aa/5B\_COMB.pep.\*
- 3: /cgn2.6/ptodata/1/1aa/6A\_COMB.pep.\*
- 4: /cgn2.6/ptodata/1/1aa/6B\_COMB.pep.\*
- 5: /cgn2.6/ptodata/1/1aa/PTCUS\_COMB.pep.\*
- 6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	4	US-09-326-203A-2
2	2309.5	83.3	503	4	US-09-593-359-4
3	1717	62.0	341	4	US-09-593-359-2
4	791.5	28.6	488	3	US-09-165-042-1
5	773.5	27.9	500	4	US-09-326-203A-17
6	770.5	27.8	498	4	US-09-103-754A-5
7	724.5	26.1	496	4	US-09-326-203A-18
8	696.5	25.1	386	4	US-09-103-754A-4
9	413	14.9	409	4	US-09-326-203A-23
10	396	14.3	550	1	US-08-121-057-4
11	396	14.3	550	2	US-08-509-187D-4
12	396	14.3	550	2	US-09-121-396-4
13	396	14.3	550	5	US-09-326-203A-4
14	375	13.5	525	4	US-09-328-857A-3
15	361	13.0	432	4	US-09-326-203A-25
16	357.5	12.9	522	3	US-09-165-042-3
17	351	12.7	502	4	US-09-328-857A-4
18	298.5	10.8	429	3	US-09-326-203A-24
19	190	6.9	51	3	US-09-165-042-9
20	153.5	5.5	52	3	US-09-165-042-6
21	147	5.3	53	3	US-09-165-042-8
22	143.5	5.2	412	4	US-09-134-001C-4885
23	143	5.2	53	3	US-09-165-042-7
24	141	5.1	52	3	US-09-165-042-5
25	123	4.4	250	4	US-09-326-203A-22
26	122.5	4.4	455	4	US-09-198-452A-1028
27	120	4.3	1235	1	US-08-118-101A-2

28	119	4.3	30	3	US-09-165-042-17	Sequence 17, Appl
29	112	4.0	352	4	US-09-205-815B-42	Sequence 42, Appl
30	102.5	3.7	33	3	US-09-165-042-10	Sequence 10, Appl
31	102.5	3.7	289	3	US-09-184-658-63	Sequence 63, Appl
32	101	3.6	813	3	US-08-836-325-8	Sequence 8, Appl
33	100	3.6	333	4	US-09-170-496D-8	Sequence 8, Appl
34	100	3.6	333	4	US-09-170-496D-168	Sequence 168, App
35	98	3.5	320	4	US-09-439-313-527	Sequence 527, App
36	97.5	3.5	461	2	US-08-463-587A-26	Sequence 26, Appl
37	97.5	3.5	461	3	US-08-463-667A-4	Sequence 4, Appl
38	97.5	3.5	461	3	US-08-923-854-26	Sequence 26, Appl
39	97.5	3.5	461	5	PCT-US91-09133-27	Sequence 27, Appl
40	97.5	3.5	975	4	US-09-328-352-4764	Sequence 4764, Ap
41	96	3.5	2237	1	US-08-455-543A-48	Sequence 48, Appl
42	96	3.5	2237	2	US-08-223-305C-48	Sequence 48, Appl
43	96	3.5	2339	1	US-08-455-543A-47	Sequence 47, Appl
44	96	3.5	2339	2	US-08-223-305C-47	Sequence 47, Appl
45	95.5	3.4	384	3	US-09-071-434-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-326-203A-2

; Sequence 2, Application US/09326203A

; Patent No. 6444876

; GENERAL INFORMATION:

; APPLICANT: Lassar, Mike

; APPLICANT: Ruesinsky, Diane

; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic

; FILE REFERENCE: 17045/00/WO

; CURRENT APPLICATION NUMBER: US/09/326, 203A

; CURRENT FILING DATE: 1999-06-04

; PRIOR APPLICATION NUMBER: 60/088,143

; PRIOR FILING DATE: 1998-06-05

; PRIOR APPLICATION NUMBER: 60/108,389

; PRIOR FILING DATE: 1998-11-12

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 520

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-326-203A-2

Query Match 100.0%; Score 2771; DB 4; Length 520;  
Best Local Similarity 100.0%; Pred. No. 2e-265;  
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAILDSAGVTTTENGGEFVLDRLRRKRS	DSNGLLSGSDNNPSDDVGPADYR	60
Db	1	MAILDSAGVTTTENGGEFVLDRLRRKRS	DSNGLLSGSDNNPSDDVGPADYR	60
Qy	61	DRIDSVDNDAGCTANLAGDNNGGDNGG	GGGGRGNADATFYRSPVAHRRARES	120
Db	61	DRIDSVDNDAGCTANLAGDNNGGDNGG	GGGGRGNADATFYRSPVAHRRARES	120
Qy	121	PLSSDAIFKQSHAGLNCVVLIAVNSRLI	ENLMKYGLWLTDFWFSRSLRDLPLFM	180
Db	121	PLSSDAIFKQSHAGLNCVVLIAVNSRLI	ENLMKYGLWLTDFWFSRSLRDLPLFM	180
Qy	181	CCISLSIFPLAAFTVEKLVQKYISEPVV	IFHIIITMTVEVLPVVTURCDSAFLSGVT	240
Db	181	CCISLSIFPLAAFTVEKLVQKYISEPVV	IFHIIITMTVEVLPVVTURCDSAFLSGVT	240
Qy	241	LMLLTCIVWLKLVYAHTSYDIRSLANA	ADKANPEVSYVSLKSLAYFMVAPTLCPQSY	300
Db	241	LMLLTCIVWLKLVYAHTSYDIRSLANA	ADKANPEVSYVSLKSLAYFMVAPTLCPQSY	300
Qy	301	PRSACIRKGWARQAKLVIFTFGFWGFI	EQVINPVRNKKPLKGLDLYAIVRLKLSV	360
Db	301	PRSACIRKGWARQAKLVIFTFGFWGFI	EQVINPVRNKKPLKGLDLYAIVRLKLSV	360





```
RESULT 12
US-10-157-855-18
; Sequence 18, Application US/10157855
; Publication No. US20020170091A1
; GENERAL INFORMATION:
; APPLICANT: Lassner, Michael W.
; APPLICANT: Ruzinsky, Diane M.
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE REFERENCE: 16516.158
; CURRENT APPLICATION NUMBER: US/10/157,855
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/326,203
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 18
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-157-855-18

Query Match      26.1%; Score 724.5; DB 14; Length 496;
Best Local Similarity 31.5%; Pred. No. 2.4e-59;
Matches 166; Conservative 98; Mismatches 166; Indels 97; Gaps 14;

QY 27 RRRKRSRSSGGLLSGSDNNSPSDVGDAPADVRDRIDSVVNDDAQGTANLAGDNNGGD 86
Db 7 RRRQPSSTSGSLASSR-----RSSEAQ 30

QY 87 NNGGGRGGEGGNADAFYTPSPVPAHRRARESPSSDALKQSHA-----GFENLCVW 141
Db 31 NNSRKSSEMRGCKEV-----VH-----TAQDSLSTSGSWTNFRGFENLSIL 75

QY 142 VLIAVNSRLIENLMKYGLIRTDWFSS-----RSLRDWLFMCICISLIFPLAAFTVE 196
Db 76 LLVLSNGRVALENVIKYLITPLQWISTFEVHHYSIWSNPLALILCSNQLISVFGME 135

QY 197 KLVQKYISEPVIFLIIITWTEVLYPYVTLRCD-SAFISGVTLMLTCTIVWLKIVSY 255
Db 136 KILRGWLGNGFAAVFTSLVIAHLTPVVVTLTHKKNPLWSVYVMGVVIEALKFTSY 195

QY 256 AHTSY---DIR-----SLANAADKA-----NPEVSY-----VSLKSLA 286
Db 196 GHVNTWARDARKTELKTQVTDLAKTCDPKQFWDLKDLSMHQMAAQAQYPAINTLSNIY 255

QY 287 YFMVAPTCYQPSYPSRACIRKGVARQAKLIVFTGPMGTIEQYINPIVRNSKHPLKG 346
Db 256 YFMAAPTLCYEFKPRLLRIRKHFKLTVELIFLSFLIALVQVQVVTVRNSMKPL-S 314

QY 347 DLLYA--IERVKLSVPNLYWLCNFYCFHLMNLIAELLCFGDREYFKDWNNAKSVGD 404
Db 315 EMEYSRCLERLLKLAIPNHLIWLFFYTFHFSFLNLAELLRFADREFYRDFWNAETIGY 374

QY 405 YWRMNMVPHKMYRHHYFPCRLSKIPKTLIIIAFLYSVAFHELCIAVPCRLKLAFL 464
Db 375 FWKGNWIPVHFAVRHHYISPMRNNFSKMSAFFVFFVSFAFFHELVSVPLKIPRLWSY 434

QY 465 GIMEQVPLVFTNTYQERFGSTGVGNMIFWFTICFGQPMCVLLYYHD 511
Db 435 GMMGQIPLSIITDKVR--GGRTGNIIVW--LSLIVGQPLAILMYGHD 478

RESULT 13
US-10-273-438-5
; Sequence 5, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
```

```
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-273-438-5

Query Match      25.1%; Score 696.5; DB 15; Length 386;
Best Local Similarity 40.9%; Pred. No. 7.4e-57;
Matches 157; Conservative 68; Mismatches 126; Indels 33; Gaps 12;

QY 150 LIIENLMKYGLIRTDWFSSRLRD---WPLFMCICISLIFPLAAFTVEKLVLYQYISE 206
Db 1 LLENLIRYGLIV-DPIQVSLFLKDPYSWAPCLVIAANVFAVAFAVQEKRLAVGALTE 59

QY 207 PVVIFLIIITWTEVLYPYVTLRCD-SAFISGVTLMLLT-CIVWLKLSYAHTSYDIRS- 264
Db 60 QAGLLHLVANIATILCFPAVVLLVESITPVGSLALMAHTILFLKLSYR---DVNSW 115

QY 265 -----LANAADKANP-----EVSY--YVSLKSLAYFMVAPTCYQPSYPSRACIRKGW 310
Db 116 CRRARAKAASAGKASSVAAPHVTSYPDNLTYRDLVLYFELFAPTCYELNFPSPRIKRF 175

QY 311 VARQPAKLVITGFMGTIEQYINPIVRNSKHPLKGLLYA--IERVKLSVPNLYWLC 368
Db 176 LLRLRLEMTFTQLQVGLIQQMVPTIIONSMKPKF-DMDYSRIIERLLKLAVPNHLIWL 234

QY 369 MEYCFEHLNLTALIELLCFGDREYFKDWNNAKSVGDYWRMNMVPHKMYRHHYFPCRLS 428
Db 235 FFYWLFSCLNAVAELMQDREYFWRMNMVPHKMYRHHYFPCRLS 294

QY 429 RIKPTLIIIAFLYSVAFHELCIAVPCRLKLAFLGIMFOVPLV-FITNYLQERFGSTV 487
Db 295 GSSKWMATGTVFLASAFHELYSVPLRMFLMAFTGMAQIPLANFVGRFGNY---- 350

QY 488 GNMIFWFTICFGQPMCVLLYYHD 511
Db 351 GNAAVW--LSLIIGQPIAVLMYVHD 373

RESULT 14
US-09-764-853-499
; Sequence 499, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PU206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 499
; LENGTH: 236
```

RESULT 11

```

US-10-273-438-10
; Sequence 10, Application US/10273438
; Publication NO. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 10
; LENGTH: 498
; TYPE: PRT
; ORGANISM: mus musculus
US-10-273-438-10

```

[illegible]

**RESULT 10**

```

US-10-273-438-7
; Sequence 7, Application US/10273438
; Publication NO. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 498
; TYPE: PRT
; ORGANISM: mus musculus
US-10-273-438-7

```

Query Match	27.8%	Score 770.5	DB 15	Length 498
Best Local Similarity	36.3%	Pred. No. 1.2e-63		
Matches 190	Conservative 83	Mismatches 179	Indels 71	Gaps 15
QY	8	GVTVTENGGEFVDLRLARRKSRSDSSNGLLLSGSDNNSPSDVGCAPADVRDRISVV	67	
DB	16	GSRSVGGSGPKVEDEVDAAVSPD-----LGAGGDAPAP-----APATRDK-----	62	
QY	68	NDDAQGTANLAGDNNGGGDDNGGGGEGRGGNADATFTYPSYPAHRRARESPSSDAI	127	
DB	63	-----DGRTSVGDGVDLRLC-----HRLQDSLFSDDSG	90	
QY	128	FKOSHAGLFWLCVVVLIVANSRLIENMKYGNLIRTDWFSSRLRD---WPLFMCCIS	184	
DB	91	F-SNYRGILNWCVVMLITLSNARLFENLIRKYGILV-DPIQVVSFLFKDPYSWPAPCVLIA	148	
QY	185	LSIFPLAAFTVEKLVLQKYISEPWVIFLIIITMTEVLVYVYVTVLRCDSAFLSGVTMLL	244	
DB	149	SNLFFVAAFQIERKLVAGLALTEQGLLHLHVNNLATAIICFPAAVALLVESTIPVGSVFALA	208	
QY	245	T-CIVMLKLVSYAHTSY-----DIRSLANAADK-----ANPEYSY-VYSLKSLAYFMV	290	
DB	209	SYSIMFLKLYSRDVLNWCQRVRKAKAVSTGKKVSGAAQQAQVSYDNLTYRDLTYFIF	268	
QY	291	APLCTQPSYPSRACIRKGWAKQFAKLVTFTGFMGFIEQYINPIYRNSKHLPLKGDLLY	350	
DB	269	APTLCEYNLNPSPRIKRELLRVLRLVLEMLFFTLQVGLIQQWVPTIHSNMKPEK-DWDY	327	
QY	351	A--IERVLKLSVNLVYWLWCMFYCFPHLWLNILAEILLCFGDREYKQDWNAKSVGDYWRM	408	
DB	328	SRRIERLLKAVPNHLIWLTLFFYWFPHSCNLVAELLOFGDREYRDWMAESYTYFWQ	387	
QY	409	WNPNVHKWVRHYTFPCRLSKIPKTLAIITAFIVSAVFELCIAVPCRLFKLWAFGLIME	468	
DB	388	WNIPVHKWIRHYKPMLRHSGSKWVARTGVFLTSAFFHEYLVSVPURMFKLWAFATAMA	447	
QY	469	QVPLVFTITNYLQBRFGSTVGNMIFWFIFCIFGQPMCVLLYYHD	511	
DB	448	QVPLAWLTVGRF-----FOGNYCNAAVW-VTLTIIGOPVAVIMVCHD	486	

420	QY	HYVFCLRSKIPKTLAIIAELVSGVSHFELCIAVPCRLFKLWAFGLMGQVPLVFTNNYL	479
241	Db	HYVFCLRKIPKVPAAIIAELVSAVSHFELCIAVPCRLFLNLMFWMGIMQVPLVFTTNFL	300
480	QY	QERFSTGVGNMIFWFIFCFIQGPQCVLLYYHDLNNRKGMS	520
301	Db	QERFSGVMGNMIFGSASCFIQPQCGLLYYHDLNNRKGMS	341

```

RESULT 8
US-10-273-438-6
; Sequence 6, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 488
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-273-438-6

```

Query Match	28.6%	Score	791.5	DB	15	Length	488
Best Local Similarity	38.0%	Pred.	No. 1.3e-65				
Matches	194	Conservative	77	Mismatches	174	Indels	65
Gaps	17						
QY	25	RLRRRSRSDSSNGLLLSGSDNNSPSDDVGCAPADVRDRIDSVVNDDAQG-TANLAGDNNG	83				
Db	8	RRRTCSRPSHCG-----CGPAAABEE-----VDAAGPDVGAAGDAPA	48				
QY	84	GDNNGGGGGGEGRGGNADATFYRPSVPAHRRARESPSSDAIFKOSHAGLFNLCVYVL	143				
Db	49	PAPNKGDDGAGVSGSHWELR-----HRLQSLFSSDSGF-SNYRGILNWCVYML	96				
QY	144	IAVNSRLIENLMKYGWLRTDFWFSRSLRD--WPLFMCCISLSFPLAAFTVEKLVL	200				
Db	97	ILSNARLFENLIKYGILV-DPTQVVSFLFKDPSHPAPCLVTAANVFAATQVERKLA	155				
QY	201	QXYISEPVWIFLHIITMEVLPVYVTLRCDSAFTSGVTMLIT-CIVWLKLVSYAHTS	259				
Db	156	VGALTEQAGLLHVNALATILCPAAVVLIVESITPVGSLALMAHTILFLKFSYR---	212				
QY	260	YDIRS-----LANAADKANP-----EVSYS--YVSLKSLAYFMVAPTLCYQSPSRSA	304				
Db	213	-DVNSWCRARAKAASAGKKAASAAAPHVYSYDNLTYRDLVYFLFAPTLCYELNFPSP	271				
QY	305	CIRKGNWAQAKLVIFTGMGBIIBQYINPIVRSNKHPLKGLDLXA--IERVLKLSVFN	362				
Db	272	RIRKRELLRILEMFLFTQVGLIQOWMPTIONSMPKPK-DWDYSRIIERLLKLAVPN	330				
QY	363	LYVWLCMFYCFHLWNLIAELLCFGDREYKDDWNAKSVGDYWRWMMNVHVKWVRHIY	422				
Db	331	HLIWLFFLFWLHSCUNVAELMQGDREYRDWNNSESITYFQNNWNIIDPHKWCIRHFY	390				
QY	423	FPCLRSKIPKTLAIIATAFLVSAVFHELCLIAVPCRLFKLWAFGLFMFQVPLV-FITNTYLOE	481				

```
Db      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |  
391 KPLRRGSKWMARTGVFLSAFHEYLVSYPRLMFRLWAEFTGMMAQIPLAWFWGRFFQG 450  
  
QY      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |  
482 RFGSTVNMIFWFETFCIFGPQCWLYYHD 511  
  
Db      | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | |  
451 NY-----GNAAVW-LSLIIIGPILVMIVYHD 475
```

```

RESULT 9
US-10-157-855-17
; Sequence 17, Application US/10157855
; Publication No. US20020170091A1
; GENERAL INFORMATION:
; APPLICANT: Lassner, Michael W.
; APPLICANT: Ruezinsky, Diane M.
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; TITLE OF INVENTION: Acid Sequences
; FILE REFERENCE: 16516.158
; CURRENT APPLICATION NUMBER: US/10/157,855
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/3326,203
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Rattus sp.
US-10-157-855-17

```

[illegible]

QY 217 TMTVEVLPVYVTLRCDLAFSGVTLMLTCTIVWLKLVSAHTSYDIRSLANAADK----- 271  
 Db 223 TTVEILFPVVLILURCDLAVLGSVTLMLFACTIVWLKLVSAHTSYDIRSLANAADK----- 282  
 QY 272 --ANPEVSYVSLKSLAYFWAPTLCYQPSYPRACIRKGVARQAKLVIFTFMGFI 329  
 Db 283 RYNWLDYADVSPKSLAYFWAPTLCYQPSYPRACIRKGVARQAKLVIFTFMGFI 342  
 QY 330 EQVINPVRNKHPLKGLDLYAIERVLKLVSPNLYVWLCMFYCFHLLWNLIAELLCFGD 389  
 Db 343 EQVINPVRNKHPLKGLDLYAIERVLKLVSPNLYVWLCMFYCFHLLWNLIAELLCFGD 402  
 QY 390 REYKDWNAKSVGDIYRWNNMPPVHKWVRHIFPCRLSKIPKTLAIIIAFLVSAFHEL 449  
 Db 403 REYKDWNAKSVGDIYRWNNMPPVHKWVRHIFPCRLSKIPKTLAIIIAFLVSAFHEL 462  
 QY 450 CIAVPCRLKFLAFLAGIMQVPLVITNTYQERF-GSTVGNMIFWTFICIFGQPMCVLLY 508  
 Db 463 CVAVPCQIFKFAFSGIMQVPLVITNTYQERF-GSTVGNMIFWTFICIFGQPMCVLLY 522  
 QY 509 YHDLNMRKGS 518  
 Db 523 YHDLNMRKAS 532

## RESULT 6

US-10-223-076-13

; Sequence 13, Application US/10223076  
 ; Publication No. US20030074695A1  
 ; GENERAL INFORMATION:

; APPLICANT: Farese, Robert V  
 ; APPLICANT: Cases, Sylvaine  
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
 ; FILE OF INVENTION: Uses Thereof

; FILE REFERENCE: UCAL-105CIP3  
 ; CURRENT APPLICATION NUMBER: US/10/223,076  
 ; CURRENT FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: 10/040,315  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 09/339,472

; PRIOR FILING DATE: 1999-06-23  
 ; PRIOR APPLICATION NUMBER: 60/107,771  
 ; PRIOR FILING DATE: 1998-11-09

; PRIOR APPLICATION NUMBER: PCT/US98/17883  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: 09/103,754

; PRIOR FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13

LENGTH: 532

TYPE: PRT

ORGANISM: Nicotiana tabacum

US-10-223-076-13

Query Match 66.5%; Score 1842.5; DB 15; Length 532;

Best Local Similarity 67.0%; Pred. No. 5.4e-164;

Matches 358; Conservative 49; Mismatches 104; Indels 23; Gaps 7;

QY 1 MAILDSAGVTVPVNGGGEFVDLD---RLRRRKRSDSSNGLLSGSDNNPSDDYGA 57  
 Db 4 MELPESVEMTTTTTSGIENLSDNHSYRRR---GNGFEAASAINSS---DANMSE 56  
 QY 58 DVDRIDS-----VVNDDAQGTANLAGDNNNGGGRGGGGRGNADAT----FTYR 108  
 Db 57 DRDVCVCSGAGLETVNERKSVGESSDVIRKEDDRNDNVANGESKSTETTPPKFAYR 116  
 QY 109 PVPARRRARESPSSDAIFKQSHAGLFNLCVVVLAVNSRLIENLMKGLIRDTDFW 168  
 Db 117 ASAPARRRARESPSSDAIFKQSHAGLFNLCVVVLAVNSRLIENLMKGLIRDTDFW 176  
 QY 169 SSRSLRDLPLFMCCISLISIFPLAFTVEKLVQKISEPVVIFLHIITMTVEVLYPVVT 228  
 Db 176 SSRSLRDLPLFMCCISLISIFPLAFTVEKLVQKISEPVVIFLHIITMTVEVLYPVVT 228

Db 177 SSKSLRDLPLFMCCISLISIFPLAFTVEKLVQKISEPVVIFLHIITMTVEVLYPVVT 236  
 QY 229 LRCDLAFSGVTLMLTCTIVWLKLVSAHTSYDIRSLANAADK---ANPEVSYVSLKSL 285  
 Db 237 LGCDLAFSGVTLMLTCTIVWLKLVSAHTSYDIRSLANAADK---ANPEVSYVSLKSL 296  
 QY 286 AYFWAPTLCYQPSYPRACIRKGVARQAKLVIFTFMGFI 345  
 Db 297 AYFWAPTLCYQPSYPRACIRKGVARQAKLVIFTFMGFI 356  
 QY 346 GDLYAIERVLKLVSPNLYVWLCMFYCFHLLWNLIAELLCFGDREYKDWNAKSVGDI 405  
 Db 357 GNLVIAIERVLKLVSPNLYVWLCMFYCFHLLWNLIAELLCFGDREYKDWNAKSVGDI 416  
 QY 406 WRNNMPPVHKWVRHIFPCRLSKIPKTLAIIIAFLVSAFHELCTAVPCRLKFLAFLG 465  
 Db 417 WRNNMPPVHKWVRHIFPCRLSKIPKTLAIIIAFLVSAFHELCTAVPCRLKFLAFLG 476  
 QY 466 IMFQVPLVITNTYQERF-GSTVGNMIFWTFICIFGQPMCVLLYHDLNMRKGS 518  
 Db 477 IMFQVPLVITNTYQERF-GSTVGNMIFWTFICIFGQPMCVLLYHDLNMRKGS 530

## RESULT 7

US-10-223-076-7

; Sequence 7, Application US/10223076  
 ; Publication No. US20030074695A1  
 ; GENERAL INFORMATION:

; APPLICANT: Farese, Robert V  
 ; APPLICANT: Cases, Sylvaine  
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and  
 ; FILE OF INVENTION: Uses Thereof

; FILE REFERENCE: UCAL-105CIP3  
 ; CURRENT APPLICATION NUMBER: US/10/223,076  
 ; CURRENT FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: 10/040,315  
 ; PRIOR FILING DATE: 2001-10-29  
 ; PRIOR APPLICATION NUMBER: 09/339,472

; PRIOR FILING DATE: 1999-06-23  
 ; PRIOR APPLICATION NUMBER: 60/107,771  
 ; PRIOR FILING DATE: 1998-11-09

; PRIOR APPLICATION NUMBER: PCT/US98/17883  
 ; PRIOR FILING DATE: 1998-08-28  
 ; PRIOR APPLICATION NUMBER: 09/103,754

; PRIOR FILING DATE: 1998-06-24  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 7

LENGTH: 341

TYPE: PRT

ORGANISM: Brassica napus

US-10-223-076-7

Query Match 62.0%; Score 1717; DB 15; Length 341;

Best Local Similarity 93.0%; Pred. No. 1.7e-152;

Matches 317; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 180 MCCISLISIFPLAFTVEKLVQKISEPVVIFLHIITMTVEVLYPVVTLCDSAFLSGV 239  
 Db 1 MCCISLISIFPLAFTVEKLVQKISEPVVIFLHIITMTVEVLYPVVTLCDSAFLSGD 60  
 QY 240 TMLLTCTIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWAPTLCYQPS 299  
 Db 61 TMLLTCTIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWAPTLCYQPS 120  
 QY 300 YPRACIRKGVARQAKLVIFTFMGFI 359  
 Db 121 YPRACIRKGVARQAKLVIFTFMGFI 180  
 QY 360 VPNLYVWLCMFYCFHLLWNLIAELLCFGDREYKDWNAKSVGDIYRWNNMPPVHKWVR 419  
 Db 181 VPNLYVWLCMFYCFHLLWNLIAELLCFGDREYKDWNAKSVGDIYRWNNMPPVHKWVR 240



Qy	121	PLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIENLMKYGLWLTIRDTWFSSRLRDWPLFM	180
Db	121	PLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIENLMKYGLWLTIRDTWFSSRLRDWPLFM	180
Qy	181	CCISLSIFPLAAAFVTEKVLVQKYLISEPVVIFLHIIITMTVELVYPVVTILRCDSAFLSGVT	240
Db	181	CCISLSIFPLAAAFVTEKVLVQKYLISEPVVIFLHIIITMTVELVYPVVTILRCDSAFLSGVT	240
Qy	241	LMLLTLCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVLSKSLAYFMVAFTLCVQPSY	300
Db	241	LMLLTLCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVLSKSLAYFMVAFTLCVQPSY	300
Qy	301	PRSAICIRKGWVARQAKLVIPTGFMGFTIEQYINPIVRNSKHLPGDGLLYATIERVLKLSV	360
Db	301	PRSAICIRKGWVARQAKLVIPTGFMGFTIEQYINPIVRNSKHLPGDGLLYATIERVLKLSV	360
Qy	361	PNLYVWLCMFCFFPHLNLINLAEELCGDREFYKDWNNAKSVGDDYWRMNMNVHKKWVRH	420
Db	361	PNLYVWLCMFCFFPHLNLINLAEELCGDREFYKDWNNAKSVGDDYWRMNMNVHKKWVRH	420
Qy	421	IYFPCLRASKTPKTLIAIIAFLVSVAFFHELCTAVPCRLEFKLNAFLGIMFQVPLVFTINYLQ	480
Db	421	IYFPCLRASKTPKTLIAIIAFLVSVAFFHELCTAVPCRLEFKLNAFLGIMFQVPLVFTINYLQ	480
Qy	481	ERFGSTVGNMIFWPFICFIFGQPMCVLLYHYHDLMNRRKGSMS	520
Db	481	ERFGSTVGNMIFWPFICFIFGQPMCVLLYHYHDLMNRRKGSMS	520

```

RESULT 2
US-10-223-076-3
; Sequence 3, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V
; APPLICANT: Cases, Sylvaine
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-223-076-3

```

```

Db 121 PLSSDAIFKQSHAGFLNLCVVVVLAVANSRLIIENLMKYGLWIRTDFESSRSLRDWPLFM 180
Qy 181 CCISLSIFPLAAFTVEKLVLQKYLISEPVWIFLHIIITWTEVLVYPVYVTLRCDSAFLSGVT 240
Db 181 CCISLSIFPLAAFTVEKLVLQKYLISEPVWIFLHIIITWTEVLVYPVYVTLRCDSAFLSGVT 240
Qy 241 LMLTCTIVWLKLVSAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFMVAPTLCYQPSY 300
Db 241 LMLTCTIVWLKLVSAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFMVAPTLCYQPSY 300
Qy 301 PRSACIRKGWVARQFAKLIVFTGPMGFIIIEQYINPIVRNSKHPKLGDLLYAIAERVLKLSV 360
Db 301 PRSACIRKGWVARQFAKLIVFTGPMGFIIIEQYINPIVRNSKHPKLGDLLYAIAERVLKLSV 360
Qy 361 PNLYVWLCMFYCFPHLMNLINLAEILLCGDRFEYKDWNAKSVGDYWRMNNPVRHKWVYRH 420
Db 361 PNLYVWLCMFYCFPHLMNLINLAEILLCGDRFEYKDWNAKSVGDYWRMNNPVRHKWVYRH 420
Qy 421 IYFPLRSKTPKTLAIIIAFLVLSAVFHELCIAVPCRFLKWAFLGIMFOVPLVITNYLQ 480
Db 421 IYFPLRSKTPKTLAIIIAFLVLSAVFHELCIAVPCRFLKWAFLGIMFOVPLVITNYLQ 480
Qy 481 ERFGSTVGNMIFWFIFCFIGQPMCVLLYYHDLNMRKGSMS 520
Db 481 ERFGSTVGNMIFWFIFCFIGQPMCVLLYYHDLNMRKGSMS 520

RESULT 3
US-10-223-076-9
; Sequence 9, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V
; APPLICANT: Cases, Sylvaine
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-223-076-9

```

Result No.	Score	Query Match	Length	DB	ID	Description
1	2771	100.0	520	14	US-10-157-855-2	Sequence 2, Appli
2	2771	100.0	520	15	US-10-223-076-3	Sequence 3, Appli
3	2309.5	83.3	508	15	US-10-223-076-9	Sequence 9, Appli
4	1880	67.8	513	15	US-10-223-076-11	Sequence 11, Appli
5	1854	66.9	534	15	US-10-223-076-15	Sequence 15, Appli
6	1842.5	66.5	532	15	US-10-223-076-13	Sequence 13, Appli
7	1717	62.0	341	15	US-10-223-076-7	Sequence 7, Appli
8	791.5	28.6	488	15	US-10-273-438-6	Sequence 6, Appli
9	773.5	27.9	500	14	US-10-157-855-17	Sequence 17, Appli
10	770.5	27.8	498	15	US-10-273-438-10	Sequence 10, Appli
11	770.5	27.8	498	15	US-10-273-438-7	Sequence 7, Appli
12	724.5	26.1	496	14	US-10-157-855-18	Sequence 18, Appli
13	696.5	25.1	386	15	US-10-273-438-5	Sequence 5, Appli
14	573.5	20.7	236	9	US-09-764-853-499	Sequence 499, Appli
15	413	14.9	409	14	US-10-157-855-23	Sequence 23, Appli



**THIS PAGE BLANK (USPTO)**

```

QY 232 DS---AFLSGVTMLLTCIWLKLVSYAHTSVDIR-----SLA 266
Db 239 KGHAFSLGASTGCVFFYSFLFKLWSYVQTNNMCROTYYQKNPRRPSITLAEKKGVL 298
QY 267 NAADRANPEVSYV-----SLKSLAYFWAPTLCYQPSYPSACIRKGVARQAKLVI 320
Db 299 NGGEE-DEDEVSKLVQYDNLTYKDLFLCAPTCYELNFPRTSRVRKRELLKLELVI 357
QY 321 FTGFMGTIEQVNPVIRNSKHPKG-DLLXAIRVLKSLVPLNLYWLCMFYCFPHLWLN 379
Db 358 GNVVYMAFQOWIIEVSNSLIPFSNMDVALATERLLKALPNHLNLCFFYLMFHSFLN 417
QY 380 ILAELLCHGDFEYKDMWNAKSGVDYWRMNMVHKKMVRHIYFPCLSKIPKTLAIIIA 439
Db 418 AVGELLNADRNFYCDMWNANNIDTFWRTWNPVHRCVRLHYIPVQMGYSSRQASTIV 477
QY 440 FLVSAVFEHLCIATVPCRLFKLWAFLGIMFOVPLVITWYLOERFSGTGNMIFWEIFCIF 499
Db 478 FLFSAVFHEYLVSVPLQIYKIWAFMGMMQIPLSAISKSIEKKLGPBMGNIIVW-ASIL 536
QY 500 GQPMCVLLYYHD 511
Db 537 GQPLCIMAYYHD 548

RESULT 15
O45245 PRELIMINARY; PRT; 496 AA.
AC O45245; O45626;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE H19N07.4 protein.
GN H19N07.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Dobson R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berka M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Fulton L.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Johnston L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Kershaw J.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; 292835; CAB07399.1;
DR EMBL; 275526; CAB07399.1; JOINED.
DR EMBL; 275526; CAB99773.1;
DR EMBL; 292835; CAB99773.1; JOINED.
DR WormPep; H19N07.4; CE16185.
DR InterPro; IPR001719; AP_endonuclease2.
DR PROSITE; PS00729; AP_NUCLEASE_F2.1; 1.
SQ SEQUENCE 496 AA; 57732 MW; 03606A2322504BE4 CRC64;

```

Query Match 26.1%; Score 724.5; DB 5; Length 496;  
 Best Local Similarity 31.5%; Pred. No. 1.7e-52;

```

Matches 166; Conservative 98; Mismatches 166; Indels 97; Gaps 14;
QY 27 RRRKSRSSNGLLILSGSDNNSPSDVGAPADVRDRIDSVYNDDAQGTANLAGDNNGGD 86
Db 7 RRRQPSETSGSLASSR-----RSSFAQ 30
QY 87 NNGGGRGCGEGRGADATFTYRPSVPAHRRARESPSLSDAIFKQSHA-----GLFNLCV 141
Db 31 NGNSSRKSSMRGPCEKV-----VH-----TAQDSLSTSGTNTFGFNLSL 75
QY 142 VLIIVNSRLIITENMKYGLINTDFWFS-----RSLRDWPLFMCCISLISFPLAFTVE 196
Db 76 LLVLSNGRVALENVIKYLITPLQWISTFEVHHYSIWSWPNLALILCSNIQILSVFGME 135
QY 197 KLVLQKYSIEPVVIFLHIIITMTVELYVPVYVTLRCD-SAFLSGVTLMLLTCTIWLKLVSY 255
Db 136 KILRGWLGNGFAAVFTYSLVIAHLTIIPVVVTLTHKWNPNLWSVVMGVVYVIEALKFISY 195
QY 256 AHTSY---DIR-----SLANAADKA-----NPEVSY-----VSLKSLA 286
Db 196 GHVNYWARDARKITELAKTQVTDLAKKTCDPKQFWDLKDLSMHQMAAQYPAHLTSLNIY 255
QY 287 YFMVAPTLCYQPSYPSACIRKGVARQAKLVIPTGFMGTIEQYINPIVNSKHPLKG 346
Db 256 YFMAAPTLCYBFFKFPRLLRIRKHFELIKRTVELIFLSFLIAALVQQWVYPTVRNSMKPL-S 314
QY 347 DLLYA--TERVLKSLVPLNLYWLCMFYCFPHLWNLILAEILLCFDGRDEYKDWNAKSGVD 404
Db 315 EMEYSRCLERLLKLAIPNHLIWLFFYTFHSHFLNLIAELLRFADREFYDFWNAETIGY 374
QY 405 YWRMNMFPVHKWVRHIYFPCLSRKIPKTLAIIIAFLYSVAFVHELCIAVPCRLPKLWAPL 464
Db 375 FWKSNIPVHREAVRHHIYSPMRNRFKMSAFFVFFVSATFEHLYSVPLKIFRLNSYY 434
QY 465 GIMFQVPLVFTINYLOERFSGTGVGMIFWIFCFIFGPMCVLLYYHD 511
Db 435 GMMGQIPLSIITDKVVR--GGRTGNIIVW--LSLIVGQPLAILMYGHD 478

```

Search completed: August 31, 2003, 04:58:58  
 Job time : 78 secs

QY 298 PSYPSACIRKQWAFQAKLVITGFMGFIIEOYINPIVRNSKHPKGLKDLIYA--IERV 355  
 Db 180 LNPFRSPRIKRLRILEMIFLQQLQVLOQMVPPTIQSMKPFK-DMYSRIERL 238  
 QY 356 LKLSVPLNVLWCMFPCFPHLWNLIALLCFGDFEYKDWNAKSVGYWRWNNPVHK 415  
 Db 239 LKLVNPHLILWIFFYFWLHSCNVAELMQGDFEYRDWNSSEYTFWQWNNIPVK 298  
 QY 416 WVRHIFPCLSKIPKTKLIIAIIAIVLSAVFHELCAVPCRLFKLWAFGLIMFQVPLV-F 474  
 Db 299 WCIRHFYKPMRLRGSSKMMARTGVLASAFHEYLVSPLRMLFRLWAFGLIMFQVPLV-F 358  
 QY 475 IYNYLQERGSGVGNWIFWIFCIGOPCMVLLIYHD 511  
 Db 359 VGRFTQGN-----GNAAVW-LSLIIGOPTAVLMVYHD 390  
 RESULT 14  
 Q96008  
 ID Q96008 PRELIMINARY; PRT: 565 AA.  
 AC Q96008;  
 DT 01-DEC-2001 (TREMELrel. 19, Created)  
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
 DE LD33852P (CG31991-PA)  
 GN MDX OR CG13273 OR CG17937 OR CG17938 OR CG31991.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkeley;  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Xu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Ananatiades P.G., Scher S.E., Li P.W., Hoskins R.A., Gallie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadenot E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Glisler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Gloeck A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy L., Muzny D.N., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasarman D.A., Weinstein K.C., Wu D., Weissenbach J.,  
 RA Williams S.M., Woodagter, Worley K.C., W. D., Yang S., Yao Q.A., Ye J.,  
 RA Zeng R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,  
 RA Evans C.A., Gocayne J.D., Ananatiades P.G., Brandon R.C., Rogers Y.,  
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,  
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,  
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,  
 RA Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,  
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,  
 RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,  
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuncio J.,  
 RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,  
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,  
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,  
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;  
 RT "Sequencing of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,  
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,  
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,  
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,  
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,  
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;  
 RT "Annotation of Drosophila melanogaster genome";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY051835; RAN93259.1; -;  
 DR EMBL; AF003652; RAN10972.1; -;  
 DR FlyBase; FBgn0004797; mdv.  
 DR InterPro; IPR001064; Crystallin.  
 DR InterPro; IPR004299; MBOAT\_fam.  
 DR Pfam; PF03062; MBOAT; 1.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 SQ SEQUENCE 565 AA; 64943 MW; 118B2AF6A385C569 CRC64;  
 Query Match 26.2%; Score 725.5; DB 5; Length 565;  
 Best Local Similarity 32.4%; Pred. No. 1.6e-52;  
 Matches 179; Conservative 96; Mismatches 202; Indels 75; Gaps 14;  
 QY 13 TENGGEFFVD-LDLRRKRKSDSNGLLSGSDNNPSDDYVGAADVDRDSDVYVND 71  
 Db 19 TNSGSSGVGIMKLRSSASATEHNLSSLRNKRSTQNLFDHGNPIDLR-QYRKVLDK-- 75  
 QY 72 OCTANLADGNNGGDNGGGRGGEGR-----GNADATFTYRPSVPAHR 115  
 Db 76 -----DENGNGTNGSEKLRRTQSVTRAETSEKKEORRAQPGRIH- 120  
 QY 116 RARESPLSDAIFKQSHAGLNLNLCVVLVIANVRLIENLKYGLWIRTDFFFSRLRD 175  
 Db 121 RPRLSLFWSGGF-TNFSGLVNWGFLLCIGGLGLENLKYGIRINPLDFWFFISGHN 179  
 QY 176 WFLFMCCISLSIFPLA----AFTVEKLVLQKVISEPVVIFLHITMTTEVLPVVTLCR 231  
 Db 180 EGEHGNALLISYLSVHLSLCLAVEKGLAMEIIAEGGLGFLQIVNVLVLCPLV-VTHL 238

```
QY 198 LVLOKYESEPVIPIHIIITTEVLYPVVYTLRCDASFLSGVTMLMLT-CIVWLKLVSYA 256
D 150 RLAVGALTEQAGLLHGVNLATILCFPAVAFALESITPPVGSVLALAMVYTLFLKLFYSR 209
QY 257 HTSYDIR-----SLANAA--DKANPEVSY--YVSLKSLAYFVMAPTLCYQPSYP 301
D 210 DVNLWCRERRRAGAKAKAAGKAANGAAQRTVSPDNLTYRDLVYFLFAPTLCYELNFP 269
QY 302 RSACIRKGVARQPAKLVIFGFMGFIIEQYINPIVRNSKHPKGLDLYA--IERVLKLS 359
D 270 RSPRIKREFLRLEMLFLTLQVGLIQQWVAIQNSMKPFK-DMYDSRIVERLLKLA 328
QY 360 VPNIYVWLCMPYCFPHLWNLITLAEELCFGDRFEYKDMWNAKSVGDYWRMNNPVMKVMVR 419
D 329 VPNIHLMIFTYLWLFHSCNLNAVAELMQFDRFYRDMWNSSEITYFWQNNWIPVHKWCIR 388
QY 420 HIYFCLRSKIPKTLAIIIAFLVSAVHFLCIAVPCRLFKLWAFELGIMFQVPLFITNYL 479
D 389 HFYKPMURGSSKWAARTAVFLASAFFHEYLVSIPLGFRFLWAFMGMAQIPLAWIVGRF 448
QY 480 QERFSTGVNMFIFCFIQOPMCVLLYYHD--LMNRK 516
D 449 ---FRGNVNAVW-LSLIIGQPVAVLMYHYHYVLRNE 483
```

## RESULT 12

```
Q8MH21 ID Q8MH21 PRELIMINARY; PRT; 489 AA.
AC Q8MH21;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Diacylglycerol acyltransferase.
GN DGAT OR DGAT1.
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Fahrenkrug S.C., Smith T.P.L., Freking B.A., Cho J., White J.,
RA Vallet J., Wise T., Rohrer G.A., Pertea G., Sultana R., Quakenbush J.,
RA Keefe J.;
RT "Porcine gene discovery by normalized cDNA-library sequencing and EST
RT cluster assembly."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Nonneman D.J., Rohrer G.A.;
RT "Linkage mapping and nucleotide variation of porcine diacylglycerol
RT acyltransferase (DGAT1).";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Nonneman D.J., Rohrer G.A.;
RT "Linkage mapping of porcine DGAT1 to a region of porcine chromosome 4
RT that contains QTL for growth and fatness."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY093657; AAM19083.1;
DR EMBL; AY116586; AAM66767.1;
DR InterPro; IPR004299; MBOAT_fam.
DR Pfam; PF03062; MBOAT; 1.
KW Acyltransferase; Transferase.
SQ SEQUENCE 489 AA; 55815 MW; 45EFB33D7F3FAACD CRC64;
```

```
Query Match 26.7%; Score 739.5; DB 6; Length 489;
Best Local Similarity 35.8%; Pred. No. 9e-54;
Matches 182; Conservative 82; Mismatches 181; Indels 63; Gaps 15;

QY 25 RLRRKRSRDSNGLLLSGSDNNSPSDVGPADVPDRIDSVNDAGGTANLADNGNG 84
D 11 RRRRTGSRPSSQSG-----SGFAAEEVRO-----VGAGGDAPTDPKDKG 52
```

```
QY 85 GDNNGGGGGGEGGRGNADATFTYRPSVPAHRRARESPSSDAIFKQSHAGLNFNCVWVLI 144
D 53 HDVSSSGHW-----DLRCH-RLQDSLFSSDSGF-SNYRGILNWCVMVLV 94
QY 145 AVNSRLIENLMKYGLWLTDFWSSRSRLD---WPLFMCISILSIFPLAFTVEKVLQ 201
D 95 LSNARLFENLIKYGILV-DPIQVVSFLFKDPYSWPAFLCVIVANFAVTAFOVEKRLAV 153
QY 202 KYISEPVVIFHIIITTEVLYPVVYTLRCDASFLSGVTMLMLT-CIVWLKLVSYAHTSY 260
D 154 GALTQAGLLIHVANLATILCFPAVAFALESITPPVGSLLALMVAITFLKLSFROVNL 213
QY 261 DIR-----SLANAADKAN-----PEVSY--YVSLKSLAYFVMAPTLCYQPSYRSC 305
D 214 WCERRRATAKAASAGKANGAAHSVSPDNLTYRDLVYFLAAPTLCYELNFSRPR 273
QY 306 IRKGWVARQPAKLVIFGFMGFIIEQYINPIVRNSKHPKGLDLYA--IERVLKLVSNL 363
D 274 IRKREFLRRLLEMLFLTLQVGLIQQWVAIQNSMKPFK-DMYDSRIVERLLKLA 332
QY 364 YVWLCMFYCFPHLWNLITLAEELCFGDRFEYKDMWNAKSVGDYWRMNNPVMKVMVRHIYF 423
D 333 LIWLIFFWLHSCNLNAVAELMQFDRFYRDMWNSSEITYFWQNNWIPVHKWCURHEK 392
QY 424 PCLRSKIPKTLAIIIAFLVSAVHFLCIAVPCRLFKLWAFELGIMFQVPLFITNYLQBRF 483
D 393 PMLRRGSSKWAARMGVFLASAFFHEYLVSIPLRFLWAFMGMAQIPLAWIVGRF---F 449
QY 484 GSTVGNMIFWEIFCFIQOPMCVLLYYHD 511
D 450 RGNYGNAVW-LSLIIGQPVAVLMYVHD 476
```

## RESULT 13

```
Q9BRH5 ID Q9BRH5 PRELIMINARY; PRT; 403 AA.
AC Q9BRH5;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Similar to diacylglycerol O-acyltransferase (Mouse) homolog
DE (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Ovary;
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006263; AAH06263.1;
KW Acyltransferase; Transferase.
FT NON_TER 1
SQ SEQUENCE 403 AA; 46662 MW; 9067C3BD78F7DAC6 CRC64;
```

```
Query Match 26.5%; Score 735.5; DB 4; Length 403;
Best Local Similarity 41.6%; Pred. No. 1.5e-53;
Matches 165; Conservative 70; Mismatches 129; Indels 33; Gaps 12;

QY 137 NLCVVVLIANSRLIENLMKYGLWLTDFWSSRSRLD---WPLFMCISILSIFPLA 193
D 5 NWCVVMLISNARLFENLIKYGILV-DPIQVVSFLFKDPYSWPAFLCVIAANFAVA 63
QY 194 TVEKLVLOKYESEPVIPIHIIITTEVLYPVVYTLRCDASFLSGVTMLMLT-CIVWLK 252
D 64 QVERKLVGALTQAGLLHGVNLATILCFPAVAFALESITPPVGSLLALMAHTLFLKL 123
QY 253 VSYAHTSYDIRS-----LANAADKANP-----EVSY--YVSLKSLAYFVMAPTLCYQ 297
D 124 FSYR----DVNSWCRARRAKAASAGKASSAAAPHTVSPDNLTYRDLVYFLAAPTLCY 179
```

```

Db 68 TSV-----GDGHWEL-----CH-RLQDSLFSSDSGF-SNY 96
QY 133 AGLFNLVVVLIIVNRLIENLMKYGLIRTDWFSSRLRD---WPLFMCCISLISIFPAAFTVEK 189
Db 97 RGLNWCVMMLLSNARLFLENLIKYGILV-DPIQVVSFLKDPYSWPALCLIVANIPAAVAFQVEK 155
QY 190 LAFTVEKVLQKYSIEPVVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVTMLLT-CIWLKLSYA 248
Db 156 VATFQIERLSLFGALTEQMLLHVNLTATILCFPAVAFLESITPVGSLTALMVTYILFLKFSYR 215
QY 249 WLKLSYAHTSY-----DIRSLANAADK-----ANPEVSY--VYSLKSLAYFMVAPTLC 295
Db 216 FLKLSYRDVNLWCRRORRKAAGVAGKVSAAQNTVSPDNLTYYEDLYFFAPLCLYELNFP 275
QY 296 YOPSYPSCASIRKGVARQAKLVITFTGFMGFIIEQYINPIVRNKHPLKGLDLYA--IERVLKLS 353
Db 276 YELNFPSPRIKRELLRVLLEMLFTQLQVGLIQOMVPAIQNSMKPKF-DMDYSRIVERLLKLA 334
QY 354 RVLKLSVPLNYVWLCMFYCFPHLWLNILAEILLCFGDREYKDWNAKSVGDYWRWNNPVP 413
Db 335 RLLKLAIVNHLIWLFFWLFHSCNLNAVAELQFGDREYRDWNAESVYTFWQNNIPV 394
QY 414 HKWVRHYIFPCLRKIPKTLAIIIAFLVSAVFHELCTIAPCRFLKWAFLGIMFQVPLV 473
Db 395 HKWCIRHYKPLRLGSKNWKWARTGVFLASAFFEHLVSIPLRMLWAFMGMAQIPLANVGRF 448
QY 474 FTNLYQERFSTVGNMIFWIFCIGFQPMCVLLYYHD 511
Db 455 WIVNRF---FQGNYGNAAVW-VTLIIGQPVAVLMVYVLD 488

```

## RESULT 10

```

Q8SQBO PRELIMINARY; PRT; 489 AA.
AC Q8SQBO;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Diacylglycerol acyltransferase 1 (EC 2.3.1.20).
GN DGAT1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21686133; PubMed=11827942;
RA Griest B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P.,
RA Camisano N., Mni M., Reid S., Simon P., Spelman R., Georges M.,
RA Snell R.;
RT "Positional Candidate Cloning of a QTL in Dairy Cattle: Identification
RT of a Missense Mutation in the Bovine DGAT1 Gene with Major Effect on
RT Milk Yield and Composition.";
RL Genome Res. 12:222-231(2002).
DR EMBL; AY065621; AAL49962.1; -.
DR InterPro; IPR004299; MBOAT_fam.
DR Pfam; PF03062; MBOAT; 1.
DR Transferrase; Acyltransferase.
SQ SEQUENCE 489 AA; 55601 MW; 4066D8C1B6743253 CRC64;

```

```

Query Match 27.4%; Score 760.5; DB 6; Length 489;
Best Local Similarity 36.8%; Pred. No. 1.5e-55;
Matches 191; Conservative 84; Mismatches 169; Indels 75; Gaps 19;

QY 27 RRRKSRSDSSNGLLSGSDNNSPD----DVGAPAD--VRDRIDSVVNDAGGTANLACD 80
Db 11 RRRRTGSRPS-----IQGGSGPAAAEVVDVAGGDAPVPRDT-----D 49
QY 81 NNGGDDNNGGGRGGEGRGADATATYRPSVPAHRRARSPSSDAIFKQSHAGLNLVCV 140
Db 50 KDGDVD-----VGSCHWD-----LRCH-RLQDSLFSSDSGF-SNYRGLNWCV 90

```

```

QY 141 VVLIIVNRLIENLMKYGLIRTDWFSSRLRD---WPLFMCCISLISIFPAAFTVEK 197
Db 91 VMLILSNARLFLENLIKYGILV-DPIQVVSFLKDPYSWPALCLIVANIPAAVAFQVEK 149
QY 198 LVLOKYSIEPVVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVTMLLT-CIWLKLSYA 256
Db 150 RIUAGALTEQAGALLHVNLTATILCFPAVAFLESITPVGSLTALMVTYILFLKFSYR 209
QY 257 HTSYDIR-----SLANAADKAN-----PEVSY--VYSLKSLAYFMVAPTLCYQPSYP 301
Db 210 DYNLWCRERRAKAKAALAGKANGAAQRTVSPDNLTYYEDLYFFAPLCLYELNFP 269
QY 302 RSACIRKGVARQAKLVITFTGFMGFIIEQYINPIVRNKHPLKGLDLYA--IERVLKLS 359
Db 270 RSPRIKRELLRVLLEMLFTQLQVGLIQOMVPAIQNSMKPKF-DMDYSRIVERLLKLA 328
QY 360 VNLVYVWLCMFYCFPHLWLNILAEILLCFGDREYKDWNAKSVGDYWRWNNPVP 419
Db 339 VPHNLIWLFFWLFHSCNLNAVAELQFGDREYKDWNAESVYTFWQNNIPV 388
QY 420 HIYFPCLRKIPKTLAIIIAFLVSAVFHELCTIAPCRFLKWAFLGIMFQVPLV 479
Db 389 HFYKPLMRGSSKWAARTAVFLASAFFEHLVSIPLRMLWAFMGMAQIPLANVGRF 448
QY 480 QERFSTVGNMIFWIFCIGFQPMCVLLYYHD--LMNRK 516
Db 449 ---FRGNYGNAAVW-VTLIIGQPVAVLMVYVLD 483

```

## RESULT 11

```

Q8MK44 PRELIMINARY; PRT; 489 AA.
AC Q8MK44;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Acyl-CoA:1,2-diacylglycerol O-transferase (EC 2.3.1.20).
GN DGAT.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Winter A., Kraemer W., Werner F.A.O., Kollers S., Kata S.,
RA Durstewitz G., Buittamp J., Womack J.E., Thaller G., Fries R.;
RT "Association of a lysine-232/alanine polymorphism in a bovine gene
RT encoding acylCoA:diacylglycerol acyltransferase (DGAT1) with variation
RT at a quantitative trait locus for milk fat content.";
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2002).
DR EMBL; AJ318490; CAC86391.1; -.
DR InterPro; IPR004299; MBOAT_fam.
DR Pfam; PF03062; MBOAT; 1.
DR Transferrase; Acyltransferase.
SQ SEQUENCE 489 AA; 55445 MW; 5C0925BDDA027053 CRC64;

```

```

Query Match 27.0%; Score 749.5; DB 6; Length 489;
Best Local Similarity 36.4%; Pred. No. 1.3e-54;
Matches 189; Conservative 85; Mismatches 170; Indels 75; Gaps 19;

```

```

QY 27 RRRKSRSDSSNGLLSGSDNNSPD----DVGAPAD--VRDRIDSVVNDAGGTANLACD 80
Db 11 RRRRTGSRPS-----IQGGSGPAAAEVVDVAGGDAPVPRDT-----D 49
QY 81 NNGGDDNNGGGRGGEGRGADATATYRPSVPAHRRARSPSSDAIFKQSHAGLNLVCV 140
Db 50 KDGDVD-----VGSCHWD-----LRCH-RLQDSLFSSDSGF-SNYRGLNWCV 90
QY 141 VVLIIVNRLIENLMKYGLIRTDWFSSRLRD---WPLFMCCISLISIFPAAFTVEK 197
Db 91 VMLILSNARLFLENLIKYGILV-DPIQVVSFLKDPYSWPALCLIVANIPAAVAFQVEK 149

```

RA Nylkiforuk C.L., Laroche A., Weselake R.J.;  
 RT "Isolation and Sequence Analysis of a Novel cDNA Encoding a Putative  
 RT Diacylglycerol Acyltransferase from a Microspora-derived Cell  
 RT Suspension Culture of *Brassica napus* L. cv Jet Neuf (Accession No.  
 RT AF155224). (PGR99-123).";  
 RL Plant Physiol. 120:1207-1207(1999).  
 DR EMBL; AF155224; AAD40861.1; -;  
 KW Acyltransferase; Transferase.  
 SQ SEQUENCE 341 AA; 39532 MW; 1A46340C49F16332 CRC64;

Query Match 62.0%; Score 1717; DB 10; Length 341;  
 Best Local Similarity 93.0%; Pred. No. 3.7e-136;  
 Matches 317; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 180 MCCISLSIFPLAFTVEKLVKQKVIISPPVIFLHIITMTLVYVYVTLRCDSAFLSGV 239  
 Db 1 MCCISLSIFPLAFTVEKLVKQKVIISPPVIFLHIITMTLVYVYVTLRCDSAFLSGD 60

Qy 240 TMLLTCTVWLKLVSYAHTSYDLSRLANADKANPEVSYVSLKSLAYFWVAPTLCYOPS 299  
 Db 61 TMLLTCTVWLKLVSYAHTNYDIRTLANSSDKANPEVSYVSLKSLAYFWMLAPTLCYOPS 120

Qy 300 YPSACIRKGVARQFAKLVFTGFMGFIETQYINPIVRNSKHPKGLDLYAIERVLKLS 359  
 Db 121 YPSFCIRKGVARQFAKLVFTGFMGFIETQYINPIVRNSKHPKGLDLYAIERVLKLS 180

Qy 360 VPVLYVWLCMFYCFHFLWNLIAELLCFDREFYKDWNAKSVGDYWRMNMMPVHKWVR 419  
 Db 181 VPVLYVWLCMFYCFHFLWNLIAELLCFDREFYKDWNAKSVGDYWRMNMMPVHKWVR 240

Qy 420 HIYPCLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFLWAFGLMFOVPLVFIINYL 479  
 Db 241 HIYPCLRIKIPKVPALIIAFLVSAVHELCIAVPCRLFLWAFGLMFOVPLVFIINFL 300

Qy 480 QERFGSVGNMIFWFIICFPGMCLVLLYVHDLNRRKGSMS 520  
 Db 301 QERFGSVGNMIFGASCFIQPMCGGLYHDLNRRKGSMS 341

## RESULT 8

Q96BB8 PRELIMINARY; PRT; 488 AA.  
 AC Q96BB8  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein (similar to diacylglycerol O-acyltransferase  
 DE homolog 1) (Mouse).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Skin;  
 RA Strausberg R.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Brain;  
 RA Strausberg R.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015762; AAH15762.1; -;  
 DR EMBL; BC023565; AAH23565.1; -;  
 DR InterPro: IPR004299; MBOAT\_fam.  
 DR Pfam: PF03062; MBOAT; 1.  
 KW Hypothetical protein; Acyltransferase; Transferase.  
 SQ SEQUENCE 488 AA; 55278 MW; 6574D5DBF15D6171 CRC64;

Query Match 28.6%; Score 791.5; DB 4; Length 488;  
 Best Local Similarity 38.0%; Pred. No. 3.8e-58;  
 Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;

## RESULT 9

Q8BH15 PRELIMINARY; PRT; 500 AA.  
 AC Q8BH15  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Diacylglycerol acyltransferase.  
 GN DEAT.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Various strains;  
 RA Harada Y., Watanabe T.K., Tanigami A.;  
 RT "rat DGAT complete cds.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB062759; BAC43739.1; -;  
 DR EMBL; AB062760; BAC43740.1; -;  
 DR EMBL; AB062761; BAC43741.1; -;  
 DR EMBL; AB062762; BAC43742.1; -;  
 DR EMBL; AB062763; BAC43743.1; -;  
 KW Acyltransferase; Transferase.  
 SQ SEQUENCE 500 AA; 57086 MW; 1A8592E7762AFFBE CRC64;

Query Match 27.9%; Score 773.5; DB 11; Length 500;  
 Best Local Similarity 37.5%; Pred. No. 1.3e-56;  
 Matches 194; Conservative 86; Mismatches 167; Indels 71; Gaps 18;

Qy 14 ENGSGFEVDLRLRRRSRSDSSNGL-LLSGSDNNSPSDVGVAPADVRDRIDSVVNDDAQ 72  
 Db 22 QGSGPMDDEEVR-----DAAVGPDLGAGGAPAPA-PVPAPAHTRDK-----DRQ 67  
 Qy 73 GTANLAGDNNGGDNNGGGGGGGGGNADATFTYRPSVPAHRRARESPLSSDAIFQSH 132

OS *Perilla frutescens*.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Lamiales; Lamiaceae; Nepetoideae; Elsholtziace;  
 OC Perilla.  
 OX NCBI\_TaxID=48386;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN=cv. Okdong;  
 RA Hwang S., Hwang Y.;  
 RT "Isolation of *Perilla frutescens* diacylglycerol acyltransferase  
 RT cDNA".  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF298815; AG23696.1; --  
 KW Acyltransferase; Transferase.  
 SQ SEQUENCE 534 AA; 61205 MW; 8FFC173E06E5BB70 CRC64;

Query Match 66.9%; Score 1854; DB 10; Length 534;  
 Best Local Similarity 65.3%; Pred. No. 1.9e-147;  
 Matches 359; Conservative 49; Mismatches 92; Indels 50; Gaps 9;

QY 1 MAILDSAGVTVT---ENGGEFVDLRLRRKRSRSDSSNGLLSGDNNSPSDVDGAP 56  
 ||||| : : : : : ||| : : : : : ||| : : : : :  
 Db 1 MAILDSPEILDTSADNGAAHHTL--RRQARSVPP-----LLDSNSL----- 47  
 ||||| : : : : : ||| : : : : : ||| : : : : :  
 QY 57 ADVDRIDSVND--DAQGTANLAGDNNNG-----GDNNG-----GRRGGE 96  
 ||||| : : : : : ||| : : : : : ||| : : : : :  
 Db 48 -----EAESAINDSENVNRDANLIENLRGAVSENEKQESYKESGARVKENGETSGN 102  
 ||||| : : : : : ||| : : : : : ||| : : : : :  
 QY 97 GRGNADATYTPSPPAHRRARESPLSSDAIFKQSHAGLFNLCVVVLIIVNSRLIENLM 156  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 103 GTDVMVAKETFPAPAHRRKRESPLSSDAIFKQSHAGLFNLCVVVLIIVNSRLIENLM 162  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 157 KYGWLIRDFWFSRSLRDWPLFMCCISLSIFPLAFTVEKLVLQYIPEVVFILHII 216  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 163 KYGWLKSGFWFSRSLRDWPLFMCCISLPPVLAFLVEKLVLKLYIPEWVAVFLHVI 222  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 217 TWEVLVYPVYVILRDSAPLSVGLTLMCTIWLKLVSVYAHTSYDIRSLANAADK----- 271  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 223 TWEVLFPVYVILRDSAPLSVGLTLMCTIWLKLVSVYAHTSYDIRSLANAADK----- 271  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 272 --ANPEVSVLSKSLAYFMWAPTLCYQSPRSACIRKGWVAROPAKLVIETGFMGFI 329  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 283 RWNLDYADVSEKSLAYFMWAPTLCYQSPRTACIRKGWVRQIKLVIETGLMGFI 342  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 330 EQYINPVRNSKHPKGLDLYAIERYKLKSVNLYVWLCMFYCFPHLNLILAEILCFGD 389  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 343 EQYINPVRNSKHPKGLDLYAIERYKLKSVNLYVWLCMFYCFPHLNLILAEILCFGD 402  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 390 REFYKDWNAKSGVDYWRMNPVHKWVRHYIYFPCRLSKIPKTLAIIIAFLVSAVFHEL 449  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 403 REFYKDWNAKSGVDYWRMNPVHKWVRHYIYFPCRLSKIPKTLAIIIAFLVSAVFHEL 462  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 450 CIAPVPCRLFKLWAFGLIMFQVPLVFTNYLQERF--GSTVGNMIFWFCIFGQPMCVLLY 508  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 463 CVAVPCIFKFWAFSGIMLQVPLIVTNYLQERKNSVMGMFMFCFFCIFGQPMCVLLY 522  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 509 YHDLNRKGS 518  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 523 YHDLNRKAS 532  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 RESULT 6  
 Q9SEG9 PRELIMINARY; PRT; 532 AA.  
 ID Q9SEG9  
 AC Q9SEG9  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Diacylglycerol acylCoA acyltransferase.  
 GN DAGAT.  
 OS Nicotiana tabacum (Common tobacco).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.  
 OX NCBI\_TaxID=4097;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN=cv. xanthi SH6; TISSUE=Callus;  
 RX MEDLINE=20069349; PubMed=10601854;  
 RA Bouvier-Nave P., Benveniste P., Oelkers P., Sturley S.L., Schaller H.;  
 RT "Expression in yeast and tobacco of plant cDNAs encoding acyl  
 RT CoA:diacylglycerol acyltransferase".  
 RL Eur. J. Biochem. 267:85-96(2000).  
 DR EMBL; AF129003; AAF19345.1; --  
 KW Acyltransferase; Transferase.  
 SQ SEQUENCE 532 AA; 60867 MW; C9D316E7A8799310 CRC64;

Query Match 66.5%; Score 1842.5; DB 10; Length 533;  
 Best Local Similarity 67.0%; Pred. No. 1.8e-146;  
 Matches 358; Conservative 49; Mismatches 104; Indels 23; Gaps 7;

QY 1 MAILDSAGVTVTENGGEFVDLD---RLRRKRSRSDSSNGLLSGDNNSPSDVDGAPA 57  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 4 MELPESVEMTTTTTSGIENLSDLNHSVRRR---GSGFEAASAINSS---DANMSE 56  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 58 DYVRDRIDS-----VYNDDAQGTANLAGDNNNGGDDNGGGRGGEGRGNADAT-----FTYR 108  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 57 DRDVCGGGAGLETVNERSKSVGESDVIRKEDDRNDNVANGESKSTTTTTTPFKFAYR 116  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 109 PSPAHRARRRESPLSSDAIFKQSHAGLFNLCVVVLIIVNSRLIENLMYGLIRDFWF 168  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 117 ASAPAHRRKRESPLSSDAIFKQSHAGLFNLCVVVLIIVNSRLIENLMYGLIRAGWF 176  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 169 SRSRLRDWPLFMCCISLSIFPLAFTVEKLVLQYIPEVVFILHIIITMTEVLYPVYVT 228  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 177 SSKSLRDWPLFMCCISLSIFPLAFTVEKLVLQYIPEVVFILHIIITMTEVLYPVYVT 236  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 229 LKCDSAFLSGVILMLTLCIWLKLVSVYAHTSYDIRSLANAADK---ANPEVSVYVLSKL 285  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 237 LKCDSAFLSGVILMLTLCIWLKLVSVYAHTSYAHNDRQLAKSTDKDETSDGDFSDVSKSL 296  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 286 AYFMWAPTLCYQSPRSACIRKGWVAROPAKLVIETGFMGFIETQYINPVRNSKHPK 345  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 297 AYFMWAPTLCYQSPRSACIRKGWVAROPAKLVIETGFMGFIETQYINPVRNSKHPK 356  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 346 GDLXAIERYKLKSVNLYVWLCMFYCFPHLNLILAEILCFGDREYKDWNAKSGVDY 405  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 357 GNLYAIERYKLKSVNLYVWLCMFYCFPHLNLILAEILCFGDREYKDWNAKSIDY 416  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 406 WRMNPVHKWVRHYIYFPCRLSKIPKTLAIIIAFLVSAVFHELCTAVPCRLFKLWAFGL 465  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 417 WRMNPVHKWVRHYIYFPCRLSKIPKTLAIIIAFLVSAVFHELCTAVPCRLFKLWAFGL 476  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 QY 466 IMFQVPLVITNYLQERF--GSTVGNMIFWFCIFGQPMCVLLYHDLNRKGS 518  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 Db 477 IMFQVPLVITNYLQERF--GSTVGNMIFWFCIFGQPMCVLLYHDLNRKGS 530  
 ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :  
 RESULT 7  
 Q9XGV4 PRELIMINARY; PRT; 341 AA.  
 ID Q9XGV4  
 AC Q9XGV4  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Putative diacylglycerol acyltransferase (EC 2.3.1.20).  
 GN DAGAT2.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=cv. Jet Neuf;

## RESULT 3

Q9XGR5 PRELIMINARY; PRT; 503 AA.  
 AC Q9XGR5;  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Putative diacylglycerol acyltransferase (EC 2.3.1.20).  
 GN DGAT1.  
 OS Brassica napus (Rape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Brassica.  
 OX NCBI\_TaxID=3708;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cultivar Jet Neuf;  
 RX MEDLINE=20027448; PubMed=10557255;  
 RA Nykiforuk C.L., Laroche A., Weselake R.J.;  
 RT "Isolation and Characterization of a cDNA Encoding a Second Putative  
 Diacylglycerol Acyltransferase from a Microspore-derived Cell  
 Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.  
 AF164434). (P699-158).";  
 RL Plant Physiol. 121:1053-1053(1999).  
 DR EMBL; AF164434; AAD45536.1; -;  
 KW Acyltransferase; Transferase.  
 SQ SEQUENCE 503 AA; 56931 MW; 2B578A16FE0AD758 CRC64;

Query Match 83.3%; Score 2309.5; DB 10; Length 503;  
 Best Local Similarity 84.8%; Pred. No. 9e-186;  
 Matches 442; Conservative 17; Mismatches 43; Indels 19; Gaps 5;  
 QY 1 MAILDSAGVT-VTEGGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGPADV 59  
 DB 1 MAILDSGGVAVPTENG---VADLRLRRKSSDSSNGLL---SDSPSDDVGAAAE 53  
 QY 60 RDRDSVNDAAQGTANLAGNNGGNGGGRGGEGGNADATFTYRSPVPAHRRARE 119  
 DB 54 RDRVDSAAEAAQGTANLA-----GGDAETRESAGG-----DVRFTYRSPVPAHRRTRE 102  
 QY 120 SPLSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWSSRLRDWPLF 179  
 DB 103 SPLSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWSSRLRDWPLF 162  
 QY 180 MCCISLSIFPLAAFTVEKLVLYSEPVVFLHIITMTVLVYVYVTLRCDSAFLSGV 239  
 DB 163 MCCLSLSVPLAAFTVEKLVLYSEPVVFLHIITMTVLVYVYVTLRCDSAFLSGV 222  
 QY 240 TMLLTCTVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWVAPLTCYOPS 299  
 DB 223 TMLLTCTVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWVAPLTCYOPS 282  
 QY 300 YPRSAIRKGVARQFAKLVFTGFMGFIIBQYINPIVRNSKHPKLGDLVLAIERVLKLS 359  
 DB 283 YPRSAIRKGVARQFAKLVFTGFMGFIIBQYINPIVRNSKHPKLGDLVLAIERVLKLS 342  
 QY 360 VPNIYVWLCMFYCFPHLWNLTAELLCFGDREYFKDMWNKSVGDYWRMNNPVHKMVR 419  
 DB 343 VPNIYVWLCMFYCFPHLWNLTAELLCFGDREYFKDMWNKSVGDYWRMNNPVHKMVR 402  
 QY 420 HIYFPCLSKIPKPLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLMFPVPLFITNVL 479  
 DB 403 HIYFPCLSKIPKPLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLMFPVPLFITNVL 462  
 QY 480 QREFGSTVGNMIFWIFCIGFQPMCVLLYYHDLNMRKGSMS 520  
 DB 463 QREFGSTVGNMIFWIFCIGFQPMCVLLYYHDLNMRKGSMS 503

## RESULT 4

Q8RX96 PRELIMINARY; PRT; 518 AA.  
 AC Q8RX96;

DT 01-JUN-2002 (TRENBLrel. 21, Created)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Putative diacylglycerol acyltransferase.  
 OS Trophaeolum majus (Common nasturtium).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OX NCBI\_TaxID=4020;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Mietkiewska E., Pedersen K., Katavic V., Taylor D.C.;  
 RT "Characterization of a putative diacylglycerol acyltransferase mRNA  
 from Trophaeolum majus embryo.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY084052; AAM03340.1; -;  
 DR InterPro; IPR004299; MBOAT.fam.  
 DR InterPro; IPR002040; Tachykinin.  
 DR Pfam; PF03062; MBOAT; 1.  
 DR PROSITE; PS00267; TACHYKININ; 1.  
 KW Transferase; Acyltransferase.  
 SQ SEQUENCE 518 AA; 58814 MW; 401D36427284CDA CRC64;  
 Query Match 67.8%; Score 1880; DB 10; Length 518;  
 Best Local Similarity 69.8%; Pred. No. 1.2e-149;  
 Matches 370; Conservative 44; Mismatches 88; Indels 28; Gaps 10;  
 QY 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSD---SSNGLLSGSDNNSPSDDVGP 56  
 DB 1 MAVAESQNTT-TMSGHGD-SDLNFRKPKSSSVIEPSSSGFT---STNGVPA---TGHV 53  
 QY 57 ADVDRDRIDSVVNDAAQGTANLAGNNGGNGGGRGGEGGR--GNADATFTYRSPVAH 114  
 DB 54 AENRQDQGVAMENATGVSNLIG-----NGGGVVIGNEEKQVGETDIRTYRSPFAH 106  
 QY 115 RRARESPUSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWSSRLSLR 174  
 DB 107 RRARESPUSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWSSRLSLR 166  
 QY 175 DWPLFMCCISLSIFPLAAFTVEKLVLYSEPVVFLHIITMTVLVYVYVTLRCDSDA 234  
 DB 167 DWSPFMCCLTPIPLAAFTVEKLVYVORHSELVAVLHVIVSTANVLPVIVLTCDSV 226  
 QY 235 FLSGVTMLLFCVWLKLVSYAHTSYDIRSLANAADK--ANP-----EVSYVSLKSLAY 287  
 DB 227 YMSGVWLMFCINWLKLVSYAHTSSDIRTLAKSGYKGAHPNSIVSCSYDVSLSLAY 286  
 QY 288 FVWAPTLCYQPSYPRSAIRKGVARQFAKLVFTGFMGFIIEQYINPIVRNSKHPKLG 347  
 DB 287 FVWAPTLCYQPSYPRSAIRKGVARQFAKLVFTGFMGFIIEQYINPIVRNSKHPKLG 346  
 QY 348 LLYAIERVLKLVSPNLYVWLCMFYCFPHLWNLTAELLCFGDREYFKDMWNKSVGDYWR 407  
 DB 347 FLYAIERVLKLVSPNLYVWLCMFYCFPHLWNLTAELLCFGDREYFKDMWNKSVGDYWR 406  
 QY 408 MNNMPVHKMVRHLYFPCLRSKIPKPLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLM 467  
 DB 407 MNNMPVHKMVRHLYFPCLRSKIPKPLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLM 466  
 QY 468 FQVPLVFTNTYLOBERP-GSTVGNMIFWIFCIGFQPMCVLLYYHDLNMRK 516  
 DB 467 FQVPLVFTNTYLOBERP-GSTVGNMIFWIFCIGFQPMCVLLYYHDLNMRK 516  
 RESULT 5  
 Q9FUL6 PRELIMINARY; PRT; 534 AA.  
 ID Q9FUL6  
 AC Q9FUL6;  
 DT 01-MAR-2001 (TRENBLrel. 16, Created)  
 DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)  
 DE Diacylglycerol acyltransferase.  
 GN DGAT1.



```

RX MEDLINE-99313150; PubMed-10386579;
RA Hills M.J., Lu C., Hobbs D.H.;
RT "Cloning of a cDNA encoding diacylglycerol acyltransferase from
RL Arabidopsis thaliana and its functional expression.";
RL FEBS Lett. 452:145-149(1999).
RN [5]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005917; AD10144.2; -
DR EMBL; AJ238008; CAB45373.1; -
DR EMBL; AF051849; AAF19262.1; -
DR EMBL; AJ131831; CAB44774.1; -
DR EMBL; AY054480; AAK96671.1; -
DR EMBL; AC003058; AAM14875.1; -
RW Acyltransferase; Transferase.
SQ SEQUENCE 520 AA; 58985 MW; 9CD0E3E8956CEFF4 CRC64;

Query Match 100.0%; Score 2771; DB 10; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.5e-224;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTENGCGGEFVDLRLRRKRSRSDSSNGLLSGDNNSPSDDVGPADVR 60
DB 1 MAILDSAGVTVTENGCGGEFVDLRLRRKRSRSDSSNGLLSGDNNSPSDDVGPADVR 60

QY 61 DRIDSVVNDDAOGTANLAGDNGGNGGGRGGEGGNADATFYRSPVPAHRRRES 120
DB 61 DRIDSVVNDDAOGTANLAGDNGGNGGGRGGEGGNADATFYRSPVPAHRRRES 120

QY 121 PLSSDAIFKQSHAGLNLVAVNSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 121 PLSSDAIFKQSHAGLNLVAVNSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180

QY 181 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 240
DB 181 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 240

QY 241 LMLTICVWLKLVSYAHTSYDRLSANAADKANPEVSYVYVSLKSLAYFWVAPLTCYQPSY 300
DB 241 LMLTICVWLKLVSYAHTSYDRLSANAADKANPEVSYVYVSLKSLAYFWVAPLTCYQPSY 300

QY 301 PRSACIRKGWARQAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLYAIEVRLKLSV 360
DB 301 PRSACIRKGWARQAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLYAIEVRLKLSV 360

QY 361 PNLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNAKSGVDYWRMNMMPVHKWVRH 420
DB 361 PNLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNAKSGVDYWRMNMMPVHKWVRH 420

QY 421 IYFPCILRSIKPTLAIITIAFLVSAVFHELCIAVPCRLFKLWAFGLIMFQVPLVFTINYLQ 480

```

```

DB 421 IYFPCILRSIKPTLAIITIAFLVSAVFHELCIAVPCRLFKLWAFGLIMFQVPLVFTINYLQ 480
QY 481 ERFGSTVGNNIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520
DB 481 ERFGSTVGNNIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 2
Q9M4V2
ID Q9M4V2 PRELIMINARY; PRT; 501 AA.
AC Q9M4V2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative diacylglycerol acyltransferase (EC 2.3.1.20).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eumetazoa II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Commanche;
RA Brown A.P., Schierer T.P., Slabas A.R.;
RT "Characterization of a putative diacylglycerol acyltransferase cDNA
RL from Brassica napus embryo.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF251794; AAF64065.1; -
RW Acyltransferase; Transferase.
SQ SEQUENCE 501 AA; 57538 MW; 29E022B278D60822 CRC64;

Query Match 84.9%; Score 2351.5; DB 10; Length 501;
Best Local Similarity 86.0%; Pred. No. 2.6e-189;
Matches 447; Conservative 17; Mismatches 37; Indels 19; Gaps 5;

QY 1 MAILDSAGVTVTENGCGGEFVDLRLRRKRSRSDSSNGLLSGDNNSPSDDVGPADVR 60
DB 1 MEILDGGVTMTPTENGCG---ADLTLRHRKPSDSSNGLL---PDSVTYSD-----ADVR 49

QY 61 DRIDSVVNDDAOGTANLAGDNGGNGGGRGGEGGNADATFYRSPVPAHRRRES 120
DB 61 DRIDSVVNDDAOGTANLAGDNGGNGGGRGGEGGNADATFYRSPVPAHRRRES 120

QY 50 DRVDSAV-EDTQGANLAGENE-----IRESGAGGNGVDVRYRSPVPAHRRRES 101
DB 50 DRVDSAV-EDTQGANLAGENE-----IRESGAGGNGVDVRYRSPVPAHRRRES 101

QY 121 PLSSDAIFKQSHAGLNLVAVNSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 102 PLSSDAIFKQSHAGLNLVAVNSRLIENLMKYGLIRTDFFWSSRLRDWPLFM 161

QY 181 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 240
DB 162 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 221

QY 241 LMLTICVWLKLVSYAHTSYDRLSANAADKANPEVSYVYVSLKSLAYFWVAPLTCYQPSY 300
DB 222 LMLTICVWLKLVSYAHTSYDRLSANAADKANPEVSYVYVSLKSLAYFWVAPLTCYQPSY 281

QY 301 PRSACIRKGWARQAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLYAIEVRLKLSV 360
DB 282 PRSACIRKGWARQAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLYAIEVRLKLSV 341

QY 361 PNLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNAKSGVDYWRMNMMPVHKWVRH 420
DB 342 PNLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNAKSGVDYWRMNMMPVHKWVRH 401

QY 421 IYFPCILRSIKPTLAIITIAFLVSAVFHELCIAVPCRLFKLWAFGLIMFQVPLVFTINYLQ 480
DB 402 VYFPCILRRNIPKPAIILAFVSAVFHELCIAVPCRLFKLWAFGLIMFQVPLVFTINYLQ 461

QY 481 ERFGSTVGNNIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520
DB 462 ERFGSTVGNNIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 501

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 04:28:05 ; Search time 72 Seconds  
(without alignments)  
1863.713 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAIDSAGVTVTENGGEF.....QPMCVLLYYHDLNMRKSGMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL\_23:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_virus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	10 Q9SLD2	Q9SLD2 arabidopsis
2	2351.5	84.9	501	10 Q9MAV2	Q9MAV2 brassica na
3	2309.5	83.3	503	10 Q9XR5	Q9XR5 brassica na
4	1880	67.8	518	10 Q8RX96	Q8RX96 tropaeolum
5	1854	66.9	534	10 Q9FUL6	Q9FUL6 perilla fru
6	1842.5	66.5	532	10 Q9SEG9	Q9SEG9 nicotiana t
7	1717	62.0	341	10 Q9XG4	Q9XG4 brassica na
8	791.5	28.6	488	4 Q96BB8	Q96BB8 homo sapien
9	773.5	27.9	500	11 Q8BRH5	Q8BRH5 rattus norv
10	760.5	27.4	489	6 Q8SQB0	Q8SQB0 bos taurus
11	749.5	27.0	489	6 Q8MK44	Q8MK44 bos taurus
12	739.5	26.7	489	6 Q8WH21	Q8WH21 sus scrofa
13	735.5	26.5	403	4 Q9BRH5	Q9BRH5 homo sapien
14	725.5	26.2	565	5 Q96OU8	Q96OU8 drosophila
15	724.5	26.1	496	5 Q45245	Q45245 caenorhabdi
16	724.5	26.1	498	5 Q9NCE1	Q9NCE1 caenorhabdi

17	720.5	26.0	565	5 Q8ST50	Q8ST50 drosophila
18	557	20.1	653	5 Q97295	Q97295 plasmodium
19	539	19.5	242	11 Q91YB5	Q91YB5 rattus norv
20	478	17.3	401	5 Q8SR71	Q8SR71 encephalito
21	419.5	15.1	462	11 Q8C795	Q8C795 mus musculu
22	397	14.3	508	6 Q8WNN5	Q8WNN5 gorilla gor
23	397	14.3	550	4 Q8NIE4	Q8NIE4 homo sapien
24	390.5	14.1	508	6 Q8WNN4	Q8WNN4 pongo pygma
25	378.5	13.7	525	11 Q8R0Y9	Q8R0Y9 mus musculu
26	374	13.5	559	5 Q8MYW9	Q8MYW9 drosophila
27	357.5	12.9	522	4 Q96TD4	Q96TD4 homo sapien
28	350.5	12.6	524	11 Q8K1W9	Q8K1W9 rattus norv
29	340.5	12.3	292	5 Q9VHN2	Q9VHN2 drosophila
30	339	12.2	580	3 Q9C2E9	Q9C2E9 neurospora
31	329	11.9	472	3 Q9U0U8	Q9U0U8 schizosacch
32	279	10.1	467	5 Q17498	Q17498 caenorhabdi
33	165.5	6.0	305	6 Q95214	Q95214 oryctolagus
34	162	5.8	609	3 Q08929	Q08929 saccharomyc
35	149.5	5.4	455	16 Q8XK89	Q8XK89 cistridium
36	147.5	5.3	621	10 Q9LNR9	Q9LNR9 arabidopsis
37	144.5	5.2	404	16 Q8CPW1	Q8CPW1 staphylococ
38	142	5.1	458	16 Q9PH05	Q9PH05 campylobact
39	139.5	5.0	404	16 Q53662	Q53662 staphylococ
40	134.5	4.9	865	16 Q8Y619	Q8Y619 listeria mo
41	127	4.6	510	16 Q8F6E3	Q8F6E3 leptospira
42	126	4.5	495	16 Q83577	Q83577 treponema p
43	122.5	4.4	473	16 Q8F4B7	Q8F4B7 leptospira
44	121.5	4.4	450	16 Q926V2	Q926V2 chlamydia p
45	121.5	4.4	1170	15 Q9PNZ5	Q9PNZ5 campylobact

#### ALIGNMENTS

#### RESULT 1

Q9SLD2 PRELIMINARY; PRT: 520 AA.

AC Q9SLD2; Q9SF22;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE Diacylglycerol O-acyltransferase (Diacylglycerol  
 DE acyltransferase).  
 GN AT2G19450 OR DGAT OR DGAT OR F3P11.5.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TAXID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,  
 RA Fraser C.M., Venter J.C.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zou J., Wei Y., Jiao C., Selvaraj G., Taylor D.C.;  
 RT "The Arabidopsis thaliana TAG1 gene encodes for a diacylglycerol  
 RT acyltransferase."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia; TISSUE-Hypocotyl;  
 RX MEDLINE=20069349; Pubmed=10601854;  
 RA Bouvier-Nave P., Benveniste P., Oalkers P., Sturley S.L., Schaller H.;  
 RT "Expression in yeast and tobacco of plant cDNAs encoding acyl  
 RT COA:diacylglycerol acyltransferase."  
 RL Eur. J. Biochem. 267:85-96(2000).  
 RN [4]  
 RP SEQUENCE FROM N.A.

```
Query Match      10.9%; Score 301; DB 1; Length 610;
Best Local Similarity 24.8%; Pred. No. 3.1e-15;
Matches 105; Conservative 76; Mismatches 161; Indels 82; Gaps 17;

QY 156 MKYGWL-IR--TDFWSSRLRDWPLFMCCISLSIFPLAA---FTVEKLVLOKYISEPVV 209
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 195 MFLGWIARCCDYASYGSA--WN-----KLEIVQYMTDLFTIAMLDLAFMLCTFFV 246
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 210 IFLHIII-----TWTEVLYPVYVTLRCDSAFLSGVTLMLLTCTIVW 249
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 247 VFVHVLVKRIINWKWTGEVAVSIFELAFIPVTFPIYV-YYEDFNWVTRIFLFLHSVFEV 305
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 250 LKLVSYA-----HTSYDIRSLANAADKANP-----EVSYY---- 279
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 306 MKSHSFAFYNGYLWDIKQBELEYSSKQLQKYKESLPETREILQKSCDFCLFELNYOTKDN 365
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 280 -----VSLKSLAYFMVAPTLCYQPSYPSACIRKGNVARQFAKLVLTGFMGFIIBQ-YI 333
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 366 DFPNNISCSNFFMFCFLFVLYVQINYPRTSRIRWRYVLEKVC-AIIGTIFLMMVTAQFTM 424
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 334 NPI-----VRNSKHPLKGDLLYAIERVLK---SVPNLVY-WLCMFYCFPHLMNIIAELL 385
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 425 HPVAMRCIOFHTPTFGGWIPATQEFHLLFDMIPGFTVLYMLTFYMIWDALLNCVAELT 484
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 386 CFGDREFYKDWNAKSVGDYRWMMMPVHKWVRHIYFFCLRS-KIPKTLAIIIAELVSA 444
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 485 RFADRYFYGDWNCVSEFEESRIWNPVHKFLLRHVYHSSMGALHLSKQATLFTFFLSA 544
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 445 VFHELCAVPCRLKFWLFGIMFQVPLVFIINYLQERFGSTVGNMIFWFIQIFGQPMC 504
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 545 VFHEMAMFAIRFRVRGYLEFMFQLSQFVWTALNTKFLARPQLSNVVFSEGVC-SGPSII 603
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 505 VLLY 508
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
QY 604 MTLY 607
Db 11: 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
```

Search completed: August 31, 2003, 04:57:31  
Job time : 42 secs

Cercopithecinae; Cercopithecus.  
 NCBI\_TaxID=9534;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=98434590; PubMed=9756918;  
 RA Anderson R.A., Joyce C., Davis M., Reagan J.W., Clark M.,  
 RA Shelness G.S., Rudel L.L.;  
 RT "Identification of a form of acyl-CoA:cholesterol acyltransferase  
 specific to liver and intestine in nonhuman primates.";  
 RL J. Biol. Chem. 273:26747-26754(1998).  
 CC -!- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY  
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE  
 CC ACTIVITY, IT MAY ACT AS A LIGASE. MAY PROVIDE CHOLESTERYL ESTERS  
 CC FOR LIPOPROTEIN SECRETION FROM HEPATOCTES AND INTESTINAL MUCOSA.  
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol  
 CC ester.  
 CC -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -!- TISSUE SPECIFICITY: LIVER AND INTESTINE. EXPRESSION IS SIX-FOLD  
 CC GREATER IN HEPATOCTES THAN IN KUPFFER CELLS.  
 CC -!- DISEASE: ACCUMULATION OF INSOLUBLE CHOLESTEROL ESTERS IN  
 CC MACROPHAGES AND SMOOTH MUSCLE IS A CHARACTERISTIC FEATURE OF EARLY  
 CC LESIONS OF ATHEROSCLEROTIC PLAQUE.  
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF053234; AAC62929.1; --  
 CC Pfam; PF03062; MBOAT; 1.  
 CC Cholesterol transferase; Transmembrane; Endoplasmic reticulum;  
 CC KW Cholesterol metabolism.  
 CC FT TRANSMEM 124 144 POTENTIAL.  
 CC FT TRANSMEM 155 175 POTENTIAL.  
 CC FT TRANSMEM 200 220 POTENTIAL.  
 CC FT TRANSMEM 308 328 POTENTIAL.  
 CC FT TRANSMEM 340 360 POTENTIAL.  
 CC FT TRANSMEM 416 436 POTENTIAL.  
 CC FT TRANSMEM 441 461 POTENTIAL.  
 CC FT TRANSMEM 476 496 POTENTIAL.  
 CC SEQUENCE 526 AA; 60411 MW; FA5E29F48386DB43 CRC64;  
 Query Match 12.6%; Score 350.5; DB 1; Length 526;  
 Best Local Similarity 26.8%; Pred. No. 5.1e-19;  
 Matches 114; Conservative 69; Mismatches 156; Indels 87; Gaps 17;  
 QY 132 HAGLENCVVLIAVNSRLIENLMKYGLWLTDFWSSRLRDWPLFMCISISIFPLA 191  
 DB 125 HMFAGLCVFII-----STLAI-----DFIDGRLLLEFDLLI--FSFGQLP 166  
 QY 192 AFTVEKLVQYISEPVVIFHIITTEVL-----YPVVTLRCDSAFLSGVT 240  
 DB 167 LVTV-----VPMFLSTLLAPYQALRLNARPCARTWILGAGLCG--ALLAAHA 212  
 QY 241 LMLIT-----CIVMLKLVSAHTSYD--IRSLANAADKAN-----PEV 276  
 DB 213 LVLCALPVHVAVEHQLPASRCVLVEQVRELMKSYSLREAVPGTLRARREGGQAPSF 272  
 QY 277 SYVYVSLKSLAYEMVAPILCYPSYRACIRKGVARQAKLVITFGFMGFIIOYINPI 336  
 DB 273 SSYL-----YFLFCPTLIYRETPRTYPIKWNVAKNFAQALGCVLYACFILGLCPV 326  
 QY 337 VRN-SKHPKGLDLYATERVLLKSVNLYVWLCMFYCFEHLWNLIAELLCFGREFYKD 395  
 DB 327 FANMSREPFSTRAL--VLSILHLATPGIFMILLIFFAFLHCLWNAFAEMLRFGDMFYRD 384

QY 396 WNAKSVGDYWRMNMMPVHKWVRHIYFP--CLRSKIPKTLAIIAFLVSAVFHELCTAV 453  
 DB 385 WNSTSFNSYRTWVNVVHDWLYSYVQDGLWLGQAARGVAMLGVLVSAVAHEYIPC- 443  
 QY 454 PCRLFKLWAFGLGIMFQVPLVF--ITNY-LQRFSGSTVGNMIFWIFCFIGQPMCVLLYYH 510  
 DB 444 ----FVLGFFVPVILFLVIGGMLNFMHMDQHTGPAWNVLMWTML-FLGOGIQVSLYC 498  
 QY 511 DLMNRK 516  
 DB 499 EWYARR 504  
 RESULT 15  
 AREL\_YEAST  
 ID AREL\_YEAST STANDARD; PRT; 610 AA.  
 AC P25628;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Sterol-ester synthase 1).  
 GN AREL OR SAT2 OR YCR048W OR YCR48W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Grivell L.A., de Haan M., Maat M.J.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE OF 1-328 FROM N.A.  
 RA Bolotin-Fukuhara M., Buhler J.-M., Daignan-Fornier B., Doira C.,  
 RA Francques-Gaillard M.-C.;  
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RN CHARACTERIZATION.  
 RX MEDLINE=96243137; PubMed=8650549;  
 RA Yang H., Bard M., Bruner D.A., Gleeson A., Deckelbaum R.J.,  
 RA Aljinovic G., Pohl T.M., Rothstein R., Sturley S.L.;  
 RL "Sterol esterification in yeast: a two-gene process.";  
 RL Science 272:1353-1356(1996).  
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol  
 CC ester.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X59720; CAA42296.1; --  
 CC PIR; S19461; S19461.  
 CC SGD; S0000644; AREL.  
 CC GO; GO:0005783; C:endoplasmic reticulum; IDA.  
 CC GO; GO:0004772; F:sterol O-acyltransferase activity; IDA.  
 CC GO; GO:0016125; P:sterol metabolism; IMP.  
 CC Pfam; PF03062; MBOAT; 1.  
 CC Transferase; Transmembrane; Endoplasmic reticulum; Acyltransferase.  
 FT TRANSMEM 182 202 POTENTIAL.  
 FT TRANSMEM 229 249 POTENTIAL.  
 FT TRANSMEM 264 284 POTENTIAL.  
 FT TRANSMEM 371 391 POTENTIAL.  
 FT TRANSMEM 409 429 POTENTIAL.  
 FT TRANSMEM 535 555 POTENTIAL.  
 FT TRANSMEM 590 610 POTENTIAL.  
 SEQUENCE 610 AA; 71613 MW; FF72EFB9238B2205 CRC64;

Db 417 LGRRARGVAMLGVELVSAVVEHYIFC-----FVLGFFYPVLMFLVFGGLLNTMNDRH 471

QY 484 GSTVGNMIFWFIFCGQPMCVLLYYHDLNMRK 516

Db 472 TGPANWILMW-TFLPMGQIOVSLYCOEWYARR 503

RESULT 13

SOA2\_HUMAN

ID SOA2\_HUMAN STANDARD; PRT; 522 AA.

AC 075908; Q9UNR2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (Cholesterol acyltransferase 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).

GN SOAT2 OR ACAT2 OR ACAT2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

SEQUENCE FROM N.A.

RP MEDLINE=98434592; PubMed=9756920;

RX Oelkers P., Behari A., Cromley D., Billheimer J.T., Sturley S.L.;

RA "Characterization of two human genes encoding acyl coenzyme

RT A:cholesterol acyltransferase-related enzymes.";

RL J. Biol. Chem. 273:26763-26771(1998).

RN [2]

SEQUENCE FROM N.A.

RP TISSUE=Intestine;

RX MEDLINE=20428724; PubMed=10846185;

RA Chang C.C.Y., Sakashita N., Ornvold K., Lee O., Chang E.T., Dong R., Lin S., Lee C.Y.G., Strom S.C., Kashyap R., Fung J.J.,

RA Farese R.V. Jr., Patoisseau J.-F., Delhon A., Chang T.-Y.;

RT "Immunological quantitation and localization of ACAT-1 and ACAT-2 in human liver and small intestine.";

RL J. Biol. Chem. 275:28083-28092(2000).

RN [3]

SEQUENCE FROM N.A.

RX MEDLINE=21225021; PubMed=111325614;

RA Katsuren K., Tamura T., Arashiro R., Takata K., Matsura T., Nikiawa N., Ohta T.;

RT "Structure of the human acyl-CoA:cholesterol acyltransferase-2 (ACAT-2) gene and its relation to dyslipidemia.";

RL Biochim. Biophys. Acta 1531:230-240(2001).

CC -!- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE ACTIVITY, IT MAY ACT AS A LIGASE. MAY PROVIDE CHOLESTERYL ESTERS FOR LIPOPROTEIN SECRETION FROM HEPATOCYTES AND INTESTINAL MUCOSA.

CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol -> CoA + cholesterol ester.

CC -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; AF059203; AAC63998.1; -

DR EMBL; AF099031; AAC78335.2; -

DR EMBL; AF331516; AAK18275.1; -

DR EMBL; AF331502; AAK18275.1; JOINED.

DR EMBL; AF331503; AAK18275.1; JOINED.

DR EMBL; AF331504; AAK18275.1; JOINED.

DR EMBL; AF331505; AAK18275.1; JOINED.

DR EMBL; AF331506; AAK18275.1; JOINED.

DR EMBL; AF331507; AAK18275.1; JOINED.

DR EMBL; AF331508; AAK18275.1; JOINED.

DR EMBL; AF331509; AAK18275.1; JOINED.

DR EMBL; AF331510; AAK18275.1; JOINED.

DR EMBL; AF331511; AAK18275.1; JOINED.

DR EMBL; AF331512; AAK18275.1; JOINED.

DR EMBL; AF331513; AAK18275.1; JOINED.

DR EMBL; AF331514; AAK18275.1; JOINED.

DR EMBL; AF331515; AAK18275.1; JOINED.

DR Genew; HGNC:11178; SOAT2.

DR MIM; 601311; -

DR GO; GO:0008415; F:acyltransferase activity; TAS.

DR GO; GO:0008203; P:cholesterol metabolism; TAS.

DR Pfam; PF03062; MBOAT; 1.

KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;

KW Cholesterol metabolism.

FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 155 175 POTENTIAL.

FT TRANSMEM 200 220 POTENTIAL.

FT TRANSMEM 262 282 POTENTIAL.

FT TRANSMEM 304 324 POTENTIAL.

FT TRANSMEM 344 366 POTENTIAL.

FT TRANSMEM 437 457 POTENTIAL.

FT TRANSMEM 472 492 POTENTIAL.

FT CONFLICT 254 254 T -> I (IN REF. 2).

SQ SEQUENCE 522 AA; 59896 MW; EEAC2DB569FF729 CRC64;

Query Match 12.9%; Score 357.5; DB 1; Length 522;

Best Local Similarity 27.8%; Pred. No. 1.5e-19;

Matches 115; Conservative 66; Mismatches 166; Indels 67; Gaps 16;

QY 132 HAGLENLCVVLIAVNSRLIENLMKYGLWLTDFWFSRSRLDWPFLWCCLISLIFPLA 191

DB 135 HMFIAGLCVFII-----STLAI-----DFIDGRLLLEFDLLI--FSGQLPLA 166

QY 192 AFTVEKLVQKYSIEPVVIFHIIITMTVLYPVVTVLRCDSAFSLGVTMLL----- 244

DB 167 LVTWVPFSLTLLAPYQALRLWARGTQA-----TGLCALLAAHAAVLCALPVHVAVE 221

QY 245 -----TCIVWLKLVSAHTSYD--IRSLANAADKAN-----PEVSYYVSLKSLAYFM 289

DB 222 HOLPPASRCVLVFEQVRFMLKSYSPLEAVPGTLRARRGEGTQAPSFSSYL-----VFL 275

QY 290 VAPTLCYOPSPRSACIRKGVARQFAKLVFTGFMGTIIQYINPIYRN--SKHPLKGD 348

DB 276 FCPITLIYRTPRTPTPVVNNYAKNFAQALGCVLYACFVLGCLVFPVFNMSREPFSTRA 335

QY 349 LYAIEKLVKLVNLYVLMCFYCFHLMNLILAECLCFGDFEYKDWNAKSVGDYWRM 408

DB 336 L--VLSILHATLPGIFMLLIIFFAFLHCLWNAFAEMLRFGDMFYRDWNSSTFSNYRT 393

QY 409 WNPVHKWVRHIYPPCLR--SKIPKTLAIIAFLVSAVFHCLTAVPCRLFKLWAFI 466

DB 394 WNVVVDWLYSVYQDGLRLLGARARGVAMLGVLVSAVAHEYIFC-----FVLGFFYPV 448

QY 467 MEQVPLVF--ITNVL--QEREGSTVGNMIFWIFCFIFGPMCVLLYYHDLNMRK 516

DB 449 MLIDFLVGGMLNFMHMDORTG--PAWNVLMWTML--FLGGQIOVSLYCOEWYARR 500

RESULT 14

SOA2\_CERAE

ID SOA2\_CERAE STANDARD; PRT; 526 AA.

AC 077759;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (Cholesterol acyltransferase 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).

GN SOAT2 OR ACAT2.

OS Cercopithecus aethiops (Green monkey) (Grivet).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

DR GO: GO:0004772; F:sterol O-acyltransferase activity; IDA.  
 DR GO:0016125; P:sterol metabolism; IMP.  
 DR Pfam; PF03062; MBOAT; 1.  
 KW Transmembrane; Endoplasmic reticulum; Acyltransferase.  
 FT TRANSMEM 215 235 POTENTIAL.  
 FT TRANSMEM 292 312 POTENTIAL.  
 FT TRANSMEM 404 424 POTENTIAL.  
 FT TRANSMEM 442 462 POTENTIAL.  
 FT TRANSMEM 485 505 POTENTIAL.  
 FT TRANSMEM 567 587 POTENTIAL.  
 FT TRANSMEM 622 642 POTENTIAL.  
 FT CONFLICT 80 80 G -> D (IN REF. 3).  
 FT CONFLICT 184 184 G -> E (IN REF. 3).  
 FT CONFLICT 211 211 I -> L (IN REF. 3).  
 FT CONFLICT 612 612 F -> S (IN REF. 3).  
 SQ SEQUENCE 642 AA; 74022 MW; 035FC4ED9C7CD830 CRC64;

Query Match 13.78; Score 378.5; DB 1; Length 642;  
 Best Local Similarity 25.28; Pred. No. 5.1e-21;  
 Matches 133; Conservative 83; Mismatches 176; Indels 135; Gaps 19;

QY 102 DAYFTYRPSV-----PAHRRARE-----SPSSDAIF 128  
 DB 125 DMSFEHPSILDSVNEPFTKVGPLEKEIRREKELAMKRLNHRKSSPDVDSVG 184  
 QY 129 KSHAGLFNLCVV-----VLIAVNSRLIEN---LMKYGWL-----IRTFWFS-SRS 172  
 DB 185 KNDGAAPTTPVTAATSETVTVTETIISNFSGLYVAFWMAIAFGAVKALIDYYQHNGS 244  
 QY 173 LRQWPLMCCISLIFPLAFTVEKLVQYISEPVVIFLH-----IIIT 217  
 DB 245 FKDEI-----LKFWTNLTVASVDLLMYLSTFYVVGIOYLCKWGLKWTGTGIFTS 298  
 QY 218 MTEVLYPVVYTLRDSA-----FLSGVTMLLTCTIVWLKLYSA-----HTSYDI 262  
 DB 299 IYEFELFVIFYMYLTENILKHLNLSKIFLHLSLVLLMKHSPAFYNGYLWGIRKEELQFSK 358  
 QY 263 RSLANAADKAN-PEV-----SYVSLKSLAYPMVAPTL 294  
 DB 359 SALAKYKDSINDPKVIGALEKSCFESLSSQSLSQTKQFPNNISAKFFWTFMPTL 418  
 QY 295 CYQPSYPRSAIRKQWVARQFAKLVITFGPMGFIIIEQYI-NPI-----VRNSKHLPLKGD 347  
 DB 419 IYQLEYPRTEIRWSYVLEKIC-AIFGTIFLMDAQILMPVAMRALAVRNS-----E 471  
 QY 348 LLXAIRVLKL-----SVENLYV-WLCMEYCFHMLNITLAEILLCFGDREYKDWNAK 400  
 DB 472 WTGILDRLKWLGLLDIVGFIWYILDYFLWDAILNCVAELTRGDRYFYGDWNCV 531  
 QY 401 SVGDYWRMNPVHKWVRHYPFCLRS-KIPKTLAIITIAFLVSAVPELHCLIAVPCRLFK 459  
 DB 532 SWADFSRIWNPVHKFLRHHVHSSMSFKLNKSQATLMTFFLSSVVELAMTVIFKKLR 591  
 QY 460 LWAFGLTMFQVPLVFTNYLQERGSVGNWIFWFCIFGQPCMLV 506  
 DB 592 FYLFFQMLQPLVALNTKFMNRRTIIGNVIFWLGICMGPSVNCML 638

## RESULT 12

SOA2\_MOUSE  
 ID SOA2\_MOUSE STANDARD; PRT; 525 AA.  
 AC O88908;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (cholesterol acyltransferase 2) (ACAT-2).  
 GN 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=98434591; PubMed=9756919;  
 RA Cases S., Novak S., Zheng Y.-W., Myers H.M., Lear S.R., Sande E.,  
 RA Welch C.B., Lusis A.J., Spencer T.A., Krause B.R., Erickson S.K.,  
 RA Farese R.V. Jr.;  
 RT "ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase. Its  
 RT cloning, expression, and characterization.";  
 RL J. Biol. Chem. 273:26755-26764(1998).  
 CC -1- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY  
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE  
 CC ACTIVITY, IT MAY ACT AS A LIGASE. MAY PROVIDE CHOLESTERYL ESTERS  
 CC FOR LIPOPROTEIN SECRETION FROM HEPATOCYTES AND INTESTINAL MUCOSA.  
 CC -1- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol -> CoA + cholesterol  
 CC ester.  
 CC -1- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -1- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF078751; AAC64057.1;  
 DR MGD; MGI:1332226; Soat2.  
 DR Pfam; PF03062; MBOAT; 1.  
 KW Transmembrane; Acyltransferase; Transmembrane; Endoplasmic reticulum;  
 KW Cholesterol metabolism.  
 FT TRANSMEM 123 143 POTENTIAL.  
 FT TRANSMEM 154 174 POTENTIAL.  
 FT TRANSMEM 197 217 POTENTIAL.  
 FT TRANSMEM 265 285 POTENTIAL.  
 FT TRANSMEM 307 327 POTENTIAL.  
 FT TRANSMEM 350 370 POTENTIAL.  
 FT TRANSMEM 440 460 POTENTIAL.  
 FT TRANSMEM 475 495 POTENTIAL.  
 SQ SEQUENCE 525 AA; 60823 MW; 01CD52E206B5C397 CRC64;

Query Match 13.58; Score 375; DB 1; Length 525;  
 Best Local Similarity 27.48; Pred. No. 7.4e-21;  
 Matches 124; Conservative 73; Mismatches 162; Indels 94; Gaps 19;

QY 112 PAHRR---ARESPSSDAIFKQSHAGLFNLCVVVLIANSRLIENLMKYGLIRTFDFWF 168  
 DB 97 PGRKKVFVARKSLT--DELMEVQH---FRTIYHMFIAGLWFLIISTL-----AIDFID 144  
 QY 169 SSRSLRDPMLFMCICISLIFPLAFTVEKLVQYISEPVVIFLHITMTTEVLYPVVVT 228  
 DB 145 EGRMLLEFDLL--FSFQQLPALMTW-----VPMFLYTLVLPVQTLW-LMAR 189  
 QY 229 LRCDSAFLSGVTL--MLLT-----CIVWLKLYSAHTSYD- 261  
 DB 190 PRAGGAMWLGASLGCVLAAHVVLCVLPVHVSVRHELPPASRCVLYFEQVRLMKSYSF 249  
 QY 262 -----IRSLANAADKANPEVSYVSLKSLAYPMVAPTLCYQPSYPRSAIRKQWVARQ 314  
 DB 250 LRETVPGIFCVRGKGISPPSFSSYL-----YFLCPTLIYREYRTPPSIRWNYVAKN 303  
 QY 315 FAKLVITFGPMGFIIIEQYINPIVFN-SKHPLKGLDLYAIAIRVLKLSV-----PNLYWLC 368  
 DB 304 FAQVGLCLLYACFTLGLRLCVFVFNANMREPPS-----TRALLSLILHATGPGIFMLL 356  
 QY 369 MFYCFHMLNITLAEILLCFGDREYKDWNAKSYGDIWRMNPVHKWVRHYPF--CL 426  
 DB 357 IFFAFLHCLWLNAPFAELRFGDRMFYRDWNSTSFNSYRTNWNVVDHLLSYVYQDGLWL 416  
 QY 427 RSKIPKTLAIITIAFLVSAVPELHCLIAVPCRLFKLWAFGLTMFQVPLV--ITNY-LQERF 483

Db 468 LSEFFPVLFVLFMFMAFN-----FIVNDSRKK---PIWVLMWTSFLGNGVLLCFYSQ 520

RESULT 10

SOAL\_RAT

ID SOAL\_RAT STANDARD; PRT; 545 AA.

AC 070536;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase 1) (ACAT-1).

DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).

GN SOAT1 OR ACAT OR ACAT.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RP [1]

RN SEQUENCE FROM N.A.

RC STRAIN=Histar; TISSUE=Adrenal gland;

RX MEDLINE=98223432; PubMed=955010;

RA Matsuda H., Hakamata H., Kawasaki T., Sakashita N., Miyazaki A., Takahashi K., Shichiri M., Horuchi S.;

RT "Molecular cloning, functional expression and tissue distribution of rat acyl-coenzyme A:cholesterol acyltransferase.";

RL Blochim. Biophys. Acta 1391:193-203(1998).

CC -|- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY CHOLESTEROL ABSORPTION.

CC -|- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol ester.

CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -|- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; D86373; BAA25372.1; -

DR Pfam; PF03062; MBOAT; 1.

KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;

KW Cholesterol metabolism.

FT TRANSMEM 136 156 POTENTIAL.

FT TRANSMEM 316 336 POTENTIAL.

FT TRANSMEM 357 377 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.

FT TRANSMEM 493 513 POTENTIAL.

SQ SEQUENCE 545 AA; 64146 MW; 40129EF21257BBF CRC64;

Query Match 14.1%; Score 391.5; DB 1; Length 545;

Best Local Similarity 28.0%; Pred. No. 4.4e-22;

Matches 132; Conservative 78; Mismatches 151; Indels 111; Gaps 24;

QY 99 GNADATFTYRPSVPAHRRARE--SPLSS-----DAIFKQSH-AGLFNLCVWVLI 144

Db 86 GCALTFTSILEEMKNHRAKDLRAPPEQKIFISRLSDELFVDHRTIYMFALLI 145

QY 145 AV-----NSRLITE-NLMKYGV-----LIRDFWFSRSRLDPLMCCISLSI 187

Db 146 IFILSTLVVDYIDEGRLVLEFSLAYAFGFPPIVWT--W-----WAMELSTLAIPY 195

QY 188 FPLAATFVKLVQKI-----SEPVVFL-----HIITMTEVL--YPVVITLRCDSAF 235

Db 196 F-----LFQWAGYKSSHPLIYSLINGAFPLVQLGILGFIPTTYVYVAYLTLP 245

QY 236 LSGVTLLMTCTIVLWKLVSAYHTSYDIRSLANAADKAN-----PEVSYVYSLKSLAYFMA 291

Db 246 ASRFILILEQIRLVMAHSYVRENVP-RVLSAKESKSVPTVQNQL-----YFLFA 298

QY 292 PILCYOPSPRSACIRKGNWARQFAKLVIITFGMGFIIEQYINPIVRNSK-HPLKGDLLY 350

Db 299 PTLIYRDSYPRPTPTVWGVAMQFLQVFGCLFVYVIFERLCAPLFRNFKQBPFS- 354

QY 351 ATERVLKLSV-----PNLYVWLCMFYCFHLLWLNILAEILLCFGRDFEYKDMNAKSVGDY 405

Db 355 ----RVVLGVFNILPGVLMFLSFFAFHCLWNAFAEMLRFGDRMFYKDMNSTSYN 411

QY 406 WRMNPNPVHKWVRHIYFPCFL--RSKIPKTLAIITIAFLVSFVHELCIAVPCR-----L 457

Db 412 YRTWNVVHDWLYVYVYKLLWFFSKRFRPAAMLAVFALSVAVVEYALAV-CLSYFPVVL 470

QY 458 FKLMAFLGIMFQVPLVFIITNYLQERGSTVGNMI-----FWFTFCIFGQ 501

Db 471 FVLEMFPGMAFN-----FIVNDSRKR---PVWIMVRASLFLGHGVILCFYSQ 515

RESULT 11

ARE2\_YEAST

ID ARE2\_YEAST STANDARD; PRT; 642 AA.

AC P53629; Q12673;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (Sterol-ester synthase 2).

GN ARE2 OR SAT1 OR YNR019W OR N3206.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI\_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Pohl T.M.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RX MEDLINE=96243137; PubMed=8650549;

RA Yang H., Bard M., Bruner D.A., Gleeson A., Deckelbaum R.J., Aljinovic G., Pohl T.M., Rothstein R., Sturley S.L.;

RT "Sterol esterification in yeast: a two-gene process.";

RL Science 272:1353-1356(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=SNY243;

RX MEDLINE=96394550; PubMed=8798656;

RA Yu C., Kennedy N.J., Chang C.C.Y., Rothblatt J.A.;

RT "Molecular cloning and characterization of two isoforms of Saccharomyces cerevisiae acyl-CoA:sterol acyltransferase.";

RL J. Biol. Chem. 271:24157-24163(1996).

CC -|- FUNCTION: ENSURES PROBABLY MOST OF THE ACYLTRANSFERASE ACTIVITY. SUPPRESSION OF ARE2 REDUCES STEROL ESTER LEVELS TO 25% OF THE NORMAL VALUE.

CC -|- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol ester.

CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -|- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; 271634; CAA36298.1; -

DR EMBL; U51790; AAB02203.1; -

DR EMBL; U55383; AAC49441.1; -

DR PIR; S63350; S63350.

DR SGD; S0005302; ARE2.

DR GO; GO:0005783; C:endoplasmic reticulum; IDA.

QY 71 AGGTANLAGDNNGGNGGGRGGGGRGNADATFTYRPSVPAHRRARESPSSDAIFKQ 130  
 DB 118 -QGKIFI-----ARRSLDELLEVDHRTI 141  
 QY 131 SHAGFNLCVVVLIIV-----NSRLIIE-NLMKYG-LIRTDFFSSRLRDWFLMCC 182  
 DB 142 YHMFALLILFLSTLVVDYIDEGRLVLEFSLLSYAFGKFPVW-----TWIMF--- 192  
 QY 183 ISLSIFPLAAFTVEKLVQKYI-----SEPVV--IFLHIITMTVEV-----LYPVVTLR 230  
 DB 193 -----LSTFSPYFLQRMATGYSKSHPLNSLFLHGFVFWQIGILGPGYVYL- 244  
 QY 231 CDSAPLGVTLMLLTCIVWLKLVSAHT---SYDIRSLANADKAN---PEVSYVSLK 283  
 DB 245 ---AYTLPPASRFIIIFEQIRFVMAKHSFVRENPRVLSAKESKSTVPITPVNYL--- 298  
 QY 284 SLAYFWAPTLCYOPSPRSACIRKGVARQAKLIVTGMFGFTIEQYINPIVNSK-H 342  
 DB 299 ---YFLFAPTLYRDSYPRNPTVRMGYVAMQFAVFGCFYVYIFERLCAPLFRNIKE 355  
 QY 343 PLKGDLLYAIAERVLKLSV-----PNLYVWLCMFYCFHLLNLILAEELLCFGDREYKDW 397  
 DB 356 PFSA-----RVLVCFNSILPGVILFLTFFAFLHCLWNAFAEMLRFGDRMEYKDW 408  
 QY 398 NAKSYGDYWRMNMVPHKWMVRHIY--FPCLRSKIPKTLAIITIAFLVSAVHELCIAYPC 455  
 DB 409 NSTYSNYRTYRNVVVDWLYYAYKDFLWFFSKRFSAAMLAVPAVSAVVEYALAV-C 467  
 QY 456 R-----LFKLWAFIGIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIQ 501  
 DB 468 LSFFYPVLVLFMFEGMAFN-----FIVNDSRKK-----PIWNVMMWTSLFLGNGVLICFY 520

## RESULT 9

SOAL\_HUMAN  
 ID SOAL\_HUMAN STANDARD; PRT: 550 AA.  
 AC P35610;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase  
 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).  
 GN SOAT1 OR SOAT OR STAT OR ACAT1 OR ACAT OR ACAT1 OR ACAT.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC TISSUE=Macrophage;  
 RX MEDLINE=94012607; PubMed=8407899;  
 RA Chang C.-Y., Huh H.-Y., Cadigan K. M., Chang T.-Y.;  
 RT "Molecular cloning and functional expression of human acyl-coenzyme  
 A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary  
 cells.";  
 RL J. Biol. Chem. 268:20747-20755(1993).  
 RN [2]  
 RP REVISION TO 207.  
 RA Chang C.-Y., Chang T.-Y.;  
 RL Submitted (May 1999) to the SWISS-PROT data bank.  
 RN [3]  
 RP TOPOLOGY.  
 RX MEDLINE=99367457; PubMed=10438503;  
 RA Lin S., Cheng D., Liu M.-S., Chen J., Chang T.-Y.;  
 RT "Human acyl-CoA:cholesterol acyltransferase-1 in the endoplasmic  
 reticulum contains seven transmembrane domains.";  
 RL J. Biol. Chem. 274:23276-23285(1999).  
 CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL  
 ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY  
 CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE  
 ACTIVITY, IT MAY ACT AS A LIGASE.  
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol -> CoA + cholesterol

ester.  
 -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS.  
 -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 reticulum.  
 -!- DISEASE: HIGHLY ACTIVATED BY THE PRESENCE OF CHOLESTEROL.  
 -!- DISEASE: ACCUMULATION OF INSOLUBLE CHOLESTEROL ESTERS IN  
 MACROPHAGES AND SMOOTH MUSCLE IS A CHARACTERISTIC FEATURE OF  
 EARLY LESIONS OF ATHEROSCLEROTIC PLAQUE.  
 -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; L21934; AAC37532.2; -;  
 DR PIR; A59038; A48026.  
 DR Genew; HGNC:11177; SOAT1.  
 DR MIM; 102642; -;  
 DR GO; GO:0005783; C:endoplasmic reticulum; TAS.  
 DR GO; GO:0008415; F:acyltransferase activity; TAS.  
 DR GO; GO:0008203; P:cholesterol metabolism; TAS.  
 DR Pfam; PF03062; MBOAT; 1.  
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;  
 KW Cholesterol metabolism.  
 FT TRANSMEM 141 159 POTENTIAL.  
 FT TRANSMEM 320 341 POTENTIAL.  
 FT TRANSMEM 361 382 POTENTIAL.  
 FT TRANSMEM 470 490 POTENTIAL.  
 FT TRANSMEM 498 518 POTENTIAL.  
 SQ SEQUENCE 550 AA; 64762 MW; 5C6AFE525D541DEE CRC64;  
 Query Match 14.3%; Score 397; DB 1; Length 550;  
 Best Local Similarity 26.7%; Pred. No. 1.7e-22;  
 Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;  
 QY 17 GGEFVLDRLRRKSRKSSNGL-----LLSGSDNNSPSDVGPADVRIDSVVND 70  
 DB 70 GSHFDFTNLIERASLDNGCALTFVSLEGEKNHRKDLRAPPE----- 117  
 QY 71 AQGTANLAGDNNGGNGGGRGGGGRGNADATFTYRPSVPAHRRARESPSSDAIFKQ 130  
 DB 118 -QGKIFI-----ARRSLDELLEVDHRTI 141  
 QY 131 SHAGFNLCVVVLIIV-----NSRLIIE-NLMKYG-LIRTDFFSSRLRDWFLMCC 182  
 DB 142 YHMFALLILFLSTLVVDYIDEGRLVLEFSLLSYAFGKFPVW-----TWIMF--- 192  
 QY 183 ISLSIFPLAAFTVEKLVQKYI-----SEPVV--IFLHIITMTVEV-LYPVVTIR 230  
 DB 193 -----LSTFSPYFLQRMATGYSKSHPLNSLFLHGFVFWQIGILGPGYVYL- 244  
 QY 231 CDSAPLGVTLMLLTCIVWLKLVSAHT---SYDIRSLANADKAN---PEVSYVSLK 283  
 DB 245 ---AYTLPPASRFIIIFEQIRFVMAKHSFVRENPRVLSAKESKSTVPITPVNYL--- 298  
 QY 284 SLAYFWAPTLCYOPSPRSACIRKGVARQAKLIVTGMFGFTIEQYINPIVNSK-H 342  
 DB 299 ---YFLFAPTLYRDSYPRNPTVRMGYVAMQFAVFGCFYVYIFERLCAPLFRNIKE 355  
 QY 343 PLKGDLLYAIAERVLKLSV-----PNLYVWLCMFYCFHLLNLILAEELLCFGDREYKDW 397  
 DB 356 PFSA-----RVLVCFNSILPGVILFLTFFAFLHCLWNAFAEMLRFGDRMEYKDW 408  
 QY 398 NAKSYGDYWRMNMVPHKWMVRHIY--FPCLRSKIPKTLAIITIAFLVSAVHELCIAYPC 455  
 DB 409 NSTYSNYRTYRNVVVDWLYYAYKDFLWFFSKRFSAAMLAVPAVSAVVEYALAV-C 467  
 QY 456 R-----LFKLWAFIGIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIQ 501



OC Cercopithecinae; Macaca.  
 RN NCBI\_TaxID=9541;  
 RP (1)  
 RC SEQUENCE FROM N.A.  
 RX TISSUE-Adrenal gland;  
 RA MEDLINE=98434590; PubMed=9756918;  
 RA Anderson R.A., Joyce C., Davis M., Reagan J.W., Clark M.,  
 RA Shelness G.S., Rudel L.L.;  
 RT "Identification of a form of acyl-CoA:cholesterol acyltransferase  
 RT specific to liver and intestine in nonhuman primates.";  
 RL J. Biol. Chem. 273:26747-26754(1998).  
 CC -!- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY  
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE  
 CC ACTIVITY, IT MAY ACT AS A LIGASE (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol  
 CC ester.  
 CC -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, BUT MOST STRONGLY  
 CC IN THE ADRENAL GLAND. EXPRESSED MORE STRONGLY IN LIVER KUPFFER  
 CC CELLS THAN IN HEPATOCYTES.  
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF053337; AAC62931.1;  
 DR Pfam; PF03062; MBOAT; 1.  
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;  
 KW Cholesterol metabolism.  
 FT TRANSMEM 142 162 POTENTIAL.  
 FT TRANSMEM 321 341 POTENTIAL.  
 FT TRANSMEM 362 382 POTENTIAL.  
 FT TRANSMEM 470 490 POTENTIAL.  
 FT TRANSMEM 498 518 POTENTIAL.  
 FT TRANSMEM 550 AA; 64699 MW; AA655206D58C291D CRC64;  
 SQ SEQUENCE  
 Query Match 14.7%; Score 406; DB 1; Length 550;  
 Best Local Similarity 26.7%; Pred. No. 3.7e-23;  
 Matches 144; Conservative 76; Mismatches 176; Indels 144; Gaps 24;  
 QY 17 GGEFVDLRLRRKSRSDSNGL-----LLSGSDNNSPSDVGPADYRDRIDSVVND 70  
 DB 70 GSHFDDFTVNLEKASLDNGGCGALTTFSILGKNNHRAKDLRAPPE----- 117  
 QY 71 AQGTANLAGDNNGGDNGGGRGGEGGCGNADATFTYRPSVPAHRRARESPSSDAIFRQ 130  
 DB 118 -QGIKFI-----ARRSLDELLEVDHIRT 141  
 QY 131 SHAGLENLCVVLIIV-----NSRLIIE-NLMKYGW-LIRTDWFSSRLDNDWPLMCC 182  
 DB 142 YHMFALLFLTLSTLWDYIDEGRLVLEFSLSYAFGRFTVW-----TWIMF----- 192  
 QY 183 ISLSIFPLAAFTVKEKLVLOKYL-----SEPVV--IFLHIITMTBV-----LYPVVTLR 230  
 DB 193 -----LSTFSVPFLFORWATGYSKSHPLNSLHGLFLPWFOIGILGFGPTVVL- 244  
 QY 231 CDASFLSGVTLMLLCIYVWKLVSXAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283  
 DB 245 ---AVTLPPASRFIIIFEQIREVFMKAHSFVRENVPRVLSAKESKSTVPIPTVQYL--- 298  
 QY 284 SLAYFWATPLCYQSPSACIRKGVARQAKLVIFGFMGFIIEQVFNIVRNSK-H 342  
 DB 299 ----YFLFAPTLIYRDSYPRNPYRWGVYAMQAFQVFCFFYVYIYPERLCAPLFRNIQ 355  
 QY 343 PLKGDLLYAIERVLKLSV-----PNLYVWLCMFYCFPHLWNLIAELLCGDFREYKDW 397  
 DB 70 GSHFDDFTVNLEKASLDNGGCGALTTFSILGKNNHRAKDLRAPPE----- 117

DB 356 PFSA-----RVLVLCVENSILPGVLILFLFTFFAFLHCWLNFAFEMLRFGDRMFYKDW 408  
 QY 398 NAKSGDYWRMNMVPHKWMRHIY--FPLRSKIPKTLAIILIAFVLSAVFHELCAVPC 455  
 DB 409 NSTSYSTYRTWNVVVDWLYYAYKDFLWFSKRFKSAMLAAFAVSAVVHEYALAV-C 467  
 QY 456 R-----LFKLWAFGLGIMFQVPLVFTNYLQRFSGTVGNMIFW-----FIFCIFOQ 501  
 DB 468 LSFYFVPLVFLFWMFMAFN---FIVNDSRKK---PIWNVMWMTSLFLGNGVLLCLFYSQ 520

## RESULT 8

SOAL\_CERAE  
 ID SOAL\_CERAE STANDARD; PRT; 550 AA.  
 AC 077760;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase  
 DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).  
 GN SOAT1 OR ACAT1.  
 OS Cercopithecus aethiops (Green monkey) (Grivet).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
 OC Cercopithecinae; Cercopithecus.  
 OX NCBI\_TaxID=9534;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX TISSUE-Adrenal gland;  
 RX MEDLINE=98434590; PubMed=9756918;  
 RA Anderson R.A., Joyce C., Davis M., Reagan J.W., Clark M.,  
 RA Shelness G.S., Rudel L.L.;  
 RT "Identification of a form of acyl-CoA:cholesterol acyltransferase  
 RT specific to liver and intestine in nonhuman primates.";  
 RL J. Biol. Chem. 273:26747-26754(1998).  
 CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL  
 CC ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY  
 CC CHOLESTEROL ABSORPTION.  
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol  
 CC ester.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, BUT MOST STRONGLY  
 CC IN THE ADRENAL GLAND. EXPRESSED MORE STRONGLY IN LIVER KUPFFER  
 CC CELLS THAN IN HEPATOCYTES.  
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF053336; AAC62930.1;  
 DR Pfam; PF03062; MBOAT; 1.  
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;  
 KW Cholesterol metabolism.  
 FT TRANSMEM 142 162 POTENTIAL.  
 FT TRANSMEM 321 341 POTENTIAL.  
 FT TRANSMEM 362 382 POTENTIAL.  
 FT TRANSMEM 470 490 POTENTIAL.  
 FT TRANSMEM 498 518 POTENTIAL.  
 FT TRANSMEM 550 AA; 64727 MW; 3B5E4CF8DB6CC713 CRC64;  
 SQ SEQUENCE

Query Match 14.5%; Score 402; DB 1; Length 550;  
 Best Local Similarity 26.5%; Pred. No. 7.3e-23;  
 Matches 143; Conservative 76; Mismatches 177; Indels 144; Gaps 24;  
 QY 17 GGEFVDLRLRRKSRSDSNGL-----LLSGSDNNSPSDVGPADYRDRIDSVVND 70  
 DB 70 GSHFDDFTVNLEKASLDNGGCGALTTFSILGKNNHRAKDLRAPPE----- 117

FT TRANSMEM 132 152 POTENTIAL.  
 FT TRANSMEM 311 331 POTENTIAL.  
 FT TRANSMEM 352 372 POTENTIAL.  
 FT TRANSMEM 460 480 POTENTIAL.  
 FT TRANSMEM 488 508 POTENTIAL.  
 FT CONFLICT 195 195 P -> R (IN REF. 2).  
 SQ SEQUENCE 540 AA; 63739 MW; 8EF900C8BCDF73C0 CRC64;  
 Query Match 15.0%; Score 415.5; DB 1; Length 540;  
 Best Local Similarity 28.5%; Pred. No. 6.9e-24;  
 Matches 135; Conservative 76; Mismatches 149; Indels 113; Gaps 24;

QY 99 GNADATFYRPSVAHRA-----RESPLSDAIFKQSHAG----- 134  
 DB 81 GCAUITSILEEMKNNHRAKOLRAPPEQKGFISRSQSL--DELFEVDHIRTTHYHMETAL 138  
 QY 135 --LENLVCVVIIVAN--SRLIIE-NLMKYGW-LIRTFWFSSRLSDWPLFMCCLSLSTFP 189  
 DB 139 LILFVLSIVVDYIDEGRLVLEFNLLAYAFKGLPTVIW-----TWAMFSLTUSIPYF- 191  
 QY 190 LAFTVEKLVLOXY-----SEPVIPL-HIITMTEVL-----YPVYVTLRCDSAFLS 237  
 DB 192 -----LFOPMHAGYSKSSHPLIYSLVHGLLFLVQLGVLPVPTVYVL-----AYTL 238  
 QY 238 GVTMLLCTIYVWLKLVSYAHT--SYDIRSLANAADKAN-----PEVSYVYSLKSLAYPMV 290  
 DB 239 PPASRFILILQELRLIMKAHSFVRENIPRVLNAAKESKSDPLPTVNOYL-----YPLF 292  
 QY 291 APTLCYOPSYPRSACIRKGWARQFAKLVIFTFMGFTIEQYINPIVNSK-HPLKGDLL 349  
 DB 293 APTLYRDNYPRTPTVRWGYYVAMQLQVFGCLFYVYIFERLCAPLFRNIKQEPFSA--- 349  
 QY 350 YAIERVLKLSV-----PNLYVWLCMFYCFPHLMNLAELLCFGDREYFKDMWNAKSYGD 404  
 DB 350 ----RVLVLCVFNLSILPGVILILSLFFAFLHCLWLNAPAEMLRFGDRMFYKDMWNSTSYN 405  
 QY 405 YWRMNPVHKWVRHIYFPCL--RSKIPKTLAIITAFVSAVHELCIAVPCR----- 456  
 DB 406 YIRTNVNVHDMLYYYAYKDLLWFFSKRFSKAAMLAVALFSAVHEYALAI-CLSYFYPV 464  
 QY 457 LFLKWAFLGIMFQVPLVFTITNLYQERFGSTVGNMIFW-----FIFCIRGQ 501  
 DB 465 LFLVLEFGMAFN---FIVNDSKR---PIWNIMWASLFLGLILCLFYSQ 510

## RESULT 6

SOAL\_CRIGR  
 ID SOAL\_CRIGR STANDARD; PRT; 546 AA.  
 AC Q60457;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase  
 DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).  
 GN SOAT1 OR ACAT1 OR ACAT.  
 OS Cricetus griseus (Chinese hamster).  
 CC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 CC Cricetulus.  
 OX NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96278939; PubMed=8662991;  
 RA Cao G., Goldstein J.L., Brown M.S.;  
 RT "Complementation of mutation in acyl-CoA:cholesterol acyltransferase  
 RT (ACAT) fails to restore sterol regulation in ACAT-defective sterol-  
 RT resistant hamster cells."  
 RL J. Biol. Chem. 271:14642-14648(1996).  
 CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL  
 CC ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY  
 CC CHOLESTEROL ABSORPTION.  
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol -> CoA + cholesterol  
 CC ester.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic  
 CC reticulum.  
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U47320; AAC52670.1; -  
 DR Pfam; PF03062; MBOAT; 1.  
 KW Transferrase; Acyltransferase; Transmembrane; Endoplasmic reticulum;  
 KW Cholesterol metabolism.  
 FT TRANSMEM 138 158 POTENTIAL.  
 FT TRANSMEM 317 337 POTENTIAL.  
 FT TRANSMEM 358 378 POTENTIAL.  
 FT TRANSMEM 466 486 POTENTIAL.  
 FT TRANSMEM 494 514 POTENTIAL.  
 SQ SEQUENCE 546 AA; 64110 MW; 4ED6C403AAC7E65B CRC64;  
 Query Match 14.9%; Score 412.5; DB 1; Length 546;  
 Best Local Similarity 28.2%; Pred. No. 1.2e-23;  
 Matches 133; Conservative 78; Mismatches 151; Indels 109; Gaps 24;

QY 99 GNADATFYRPSVAHRAARE--SPLSS-----DAIFKQSH-----AGLEN 137  
 DB 87 GCAUITSILEEMKNNHRAKOLRAPPEKGFISRRSLDELFEVDHIRTTHYHPIGLLI 146  
 QY 138 LCVVLIIVAN-----SRLIIE-NLMKYGW-LIRTFWFSSRLSDWPLFMCCLSLIFPLA 191  
 DB 147 LFILSTLVVDYIDEGRLVLEFNLLAYAFKGLPTVIW-----TWAMFSLTUSIPYF--- 197  
 QY 192 AFTVEKLVLOXY-----ISEPVIFL-----HIITMTEVL--YPVYVTLRCDSAFLSGV 239  
 DB 198 -----LFOPMHAGYSKTSHPLIYSLHSGFFLVQLGILGFPVPTVYVL-----AYTLPP 246  
 QY 240 TMLLCTIYVWLKLVSYAHT--SYDIRSLANAADKAN-----PEVSYVYSLKSLAYFVAP 292  
 DB 247 ASRFIVILEQIRMYHKAHSFVRENIPRVLNAAKESSTVPVPTVNOYL-----YFLFAP 300  
 QY 293 TLCYOPSYPRSACIRKGWARQFAKLVIFTFMGFTIEQYINPIVNSK-HPLKGDLLYA 351  
 DB 301 TLIVRDSYPTPTVRWGYYVAMQLQVFGCLFYVYIFERLCAPLFRNIKQEPFSA----- 355  
 QY 352 IERVLKLSV-----PNLYVWLCMFYCFPHLMNLAELLCFGDREYFKDMWNAKSYGDY 406  
 DB 356 --RVLVLCVFNLSILPGVILMLFTFAFLHCLWLNAPAEMLRFGDRMFYKDMWNSTSYN 413  
 QY 407 RMWNPVHKWVRHIYFPCL--RSKIPKTLAIITAFVSAVHELCIAVPCR-----LF 458  
 DB 414 RTWNVVHDMLYYYAYKDLLWFFSKRFSKAAMLAVALFSAVHEYALAV-CLSYFYPVLF 472  
 QY 459 KLWAFGLGIMFQVPLVFTITNLYQERFGSTVGNMIFW-----FIFCIRGQ 501  
 DB 473 VLEHFGMAFN---FIVNDSKR---PIWNIMWASLFLGLHGLVILCLFYSQ 516

RESULT 7  
 SOAL\_MACFA  
 ID SOAL\_MACFA STANDARD; PRT; 550 AA.  
 AC Q77761;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase  
 DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).  
 GN SOAT1 OR ACAT1 OR ACAT.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;



DE acyltransferase).

GN DGAT1 OR DGAT.

OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RX MEDLINE=99007259; PubMed=9789033;

RA Cases S., Smith S.J., Zheng Y.-W., Myers H.M., Lear S.R., Sande E.,

RA Novak S., Collins C., Welch C.B., Lusis A.J., Erickson S.K.,

RA Farese R.V. Jr.;

RT Identification of a gene encoding an acyl CoA:diacylglycerol

RT acyltransferase, a key enzyme in triacylglycerol synthesis.\*;

RL Proc. Natl. Acad. Sci. U.S.A. 95:13018-13023(1998).

[2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Tongue;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakomoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuiki S.,

RA Hayashizaki Y.;

RT \*Functional annotation of a full-length mouse cdna collection.\*;

RL Nature 409:685-690(2001).

[3]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Udell T.B., Toshyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT \*Generation and initial analysis of more than 15,000 full-length

RT human and mouse cdna sequences.\*;

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- FUNCTION: Catalyzes the terminal and only committed step in

CC triacylglycerol synthesis by using diacylglycerol and fatty acyl

CC CoA as substrates

CC -1- CATALYTIC ACTIVITY: Acyl-CoA + 1,2-diacylglycerol -> CoA +

CC triacylglycerol.

CC -1- PATHWAY: Central role in the metabolism of cellular diacylglycerol

CC lipids.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic

CC reticulum.

CC -1- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

CC -----

DR EMBL: AF078752; AAC72917.1;

DR EMBL: AK008995; -; NOT\_ANNOTATED\_CDS.

DR EMBL: BC003717; AA03717.1; -;

DR MGD: MGI:1333825; Dgat1.

DR Pfam: PF03062; MBOAT; 1.

KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum.

FT TRANSMEM 113 133 POTENTIAL.

FT TRANSMEM 139 159 POTENTIAL.

FT TRANSMEM 175 195 POTENTIAL.

FT TRANSMEM 198 218 POTENTIAL.

FT TRANSMEM 293 313 POTENTIAL.

FT TRANSMEM 343 363 POTENTIAL.

FT TRANSMEM 417 437 POTENTIAL.

FT TRANSMEM 439 459 POTENTIAL.

FT TRANSMEM 464 484 POTENTIAL.

SQ SEQUENCE 498 AA; 56790 MW; E7B0DD6DDCF1EC2B CRC64;

Query Match 27.8%; Score 771.5; DB 1; Length 498;

Best Local Similarity 36.3%; Pred. No. 1.2e-50;

Matches 190; Conservative 84; Mismatches 178; Indels 71; Gaps 15;

QY 8 GVTVTVTGGGFEVDLRLRRKRSRSDSSNGLLSGDNNPSDDVGAPADVRDRIDSV 67

DB 16 GSRVSVGGGSGPKVEEDVRDAVSPD-----LCAGGDAPAPAP-----APAHRDK----- 62

QY 68 NDDAQTANTAGDNNGGGCGGEGRGGNADATFTYRPSVPAHRARRESPLSDAI 127

DB 63 -----DGRTSVGDGWDLR-----HRLQDSLFSSDSG 90

QY 128 FKQSHAGLNLVVVLIIVANSRLIENLMKVGILRTDFWFSRSLRD---WPLFMCCIS 184

DB 91 F-SNYRGILNWCVVVLLISNARLENLKYGILV-DPIQVVSFLKDPYSWPAPCVIIA 148

QY 185 LSIFPLAAFTVEKLVQKYSIEPVVIFLHIITMTEVLYPVVYVTLRCDSAFLSGVTMLL 244

DB 149 SNIFVVAFAQIEKRLAVGALTEQMGLLHVVNLATIIICFPAAVALLVESITPVGSVFALA 208

QY 245 T-CIVWLKLYSYAHTSY-----DIRSLANAADK-----ANPEVSY--VYSLKSLATPMV 290

DB 209 SYSIMFLKLYSYRDVNLWCQRVRVAKAVSTGKRVSGAAQAQVSPDNLTLYDYFIF 268

QY 291 APTLCYQPSYPRSAICIRKGVVARQFAKLVIPTGFMFTIEQYINPIVRNSKHPKLGDL 350

DB 269 APTLCYELNPSRSPRIKRLRLRVLEMLFTQIQVLIQWVPTTONSKMKPKF-DMDY 327

QY 351 A--IERVKLKSVNLYVWLCMFYCFHMLNLIAELLCFGDREYFKDWNNAKSYGDYWRM 408

DB 328 SRIERELLKLVNPHLIWLIFFYWFHSCNLNAVALLELQFGDREYFDWNNAESVYTFWQN 387

QY 409 WNPVHKWVRHYFPCRLSKIPKTLAIITAFLYSAVFHELCIAPVCRLEKFLWFLGIMP 468

DB 388 WNPVHKWVRHYFPCRLSKIPKTLAIITAFLYSAVFHELCIAPVCRLEKFLWFLGIMP 468

QY 469 QVPLVFTNTVQLQERFGSTVGNMIFWIFCFIGQPMCVLLVYVD 511

DB 448 QVPLAWIVGRF---FQGNYNAAVW-VTLIGQPVAVLMVYVD 486

RESULT 4

ID DGT1\_RAT

IGT1\_RAT STANDARD; PRT; 498 AA.

AC O9PRM3;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

FT TRANSMEM 453 473 POTENTIAL.
SQ SEQUENCE 488 AA; 55252 MW; 12E34BA7478ABAIF CRC64;

Query Match 28.6%; Score 791.5; DB 1; Length 488;
Best Local Similarity 38.0%; Pred. No. 3.7e-52;
Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;

QY 25 RLRRKRSRSDSSNGLLGGSDNNPSDDVAGADVPDRDRIDSVVNDDAQG-TANLAGDNG 83
  - - - - -
Db 8 RRRRTGSRPSSHG - - - - -GGPAAAE - - - - -VRDAAAGPDVGAAGDAFA 48

QY 84 GGDNGGGGGRGGRGNADATFYRSPVPAHRRARESPSSDAIFKQSHAGFLNLCVVYL 143
  - - - - -
Db 49 PAPKDDAGVSGHWELRC - - - - -HRLQSLFSSDSGF-SNTRGILNWCVVML 96

QY 144 IAVNSRLIENLMKYGLIRTDWFSSRLD - - - - -WPLEMCCISLSIFPLAFTVEKLV 200
  - - - - -
Db 97 ILSNARLENLIKYGLV-DPIQVVSFLKOPHSPAPCLVIAANVFAVAAFOVEKRLA 155

QY 201 QKYSIEPVVFIHIIITMTVEVLPVVTLCRDSAFSLGVTMLLT-CIWLKLVSAHTS 259
  - - - - -
Db 156 VQALTEQAGLLHVNATILCFPAAVLLVESITPVGSLALMAHTILFLKLSYR - - - 212

QY 260 YDTRS - - - - -LANAADKANP - - - - -EVS - - - - -YVLSKSLAYFMVAPTLCYPSYPSA 304
  - - - - -
Db 213 -DVNSWCRARAKAASAGKASSAAAPHTVSPDNLTIRDLTYFLFAPILCYELNPPSP 271

QY 305 CIRKGWVAFQKLVITFGFMGFIIEQYINPIVRNSKHLKGLDLYA - - - - -IERVILKLVSPN 362
  - - - - -
Db 272 RIRKRELLRLEMLFTQVLQVLIQOMVPTIOMSKPKF-DMDYSRIERLLKLVN 330

QY 363 LYVWLCMFCFFHLNIIAELLCFGRDFYKDWNAKSVGDYWRWNNPVHKKWVRHIY 422
  - - - - -
Db 331 HLWIFFYFLHSCILNAELMQLFGDFEYRDNWNSSEVTYFWQNNIPVHKKWIRHEY 390

QY 423 FPLCRSKIPKTLAIIITAFVSAVFHELCTAVPCRLKFLWAFGIMFQVPLV-FITVLOE 481
  - - - - -
Db 391 KPLMRGSKWMAWRTGVFLASAFFHEYLVSPLRMFLWAFGMAQIPLANEVGRFFQG 450

QY 482 RFGSTVGNMIFWIFCIGOPCMVLLYYHD 511
  - - - - -
Db 451 NY - - - - -GNAAVW-LSLIIGQPIAVLMYVHD 475

RESULT 2
DGT1_CERAE
ID DGT1_CERAE STANDARD; PRT; 491 AA.
AC Q9GMF1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Diacylglycerol O-acyltransferase 1 (EC 2.3.1.20) (Diglyceride acyltransferase).
DE DGAT1 OR DGAT.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Joyce C.W., Davis M.A., Anderson R.A., Rudel L.L.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -FUNCTION: Catalyzes the terminal and only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates.
CC -CATALYTIC ACTIVITY: Acyl-CoA + 1,2-diacylglycerol = CoA + triacylglycerol.
CC -PATHWAY: Central role in the metabolism of cellular diacylglycerol lipids.
CC -SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (By similarity).

-1- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AF236018; M60AT: 1;
DR Pfam; PF03062; M60AT: 1;
KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum.
FT TRANSMEM 107 127
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 404 424 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT TRANSMEM 456 476 POTENTIAL.
SQ SEQUENCE 491 AA; 55643 MW; BFD3683453D588DB CRC64;

Query Match 27.9%; Score 772.5; DB 1; Length 491;
Best Local Similarity 37.8%; Pred. No. 9.9e-51;
Matches 191; Conservative 80; Mismatches 179; Indels 55; Gaps 17;

QY 25 RLRRKRSRSDSSNGLLGGSDNNPSDDVAGADVPDRDRIDSVVNDDAQG-TANLAGDNGS 84
  - - - - -
Db 11 RRRRTGSRPSSHG - - - - -GGPAAAE - - - - -VRDAAAG - - - - -DMGAA 46

QY 85 GDNNGGGGGRGGRGNADATFYRSPVPAHRRARESPSSDAIFKQSHAGFLNLCVVYL 144
  - - - - -
Db 47 GDAPAPAPSKDADGAVASGHWELR - - - - -CH-RLQDSLFSSDSGF-NNYRGILNWCVVML 100

QY 145 AVNSRLIENLMKYGLIRTDWFSSRLD - - - - -WPLEMCCISLSIFPLAFTVEKLV 201
  - - - - -
Db 101 LSNARLENLIKYGLV-DPIQVVSFLKOPHSPAPCLVIAANVFAVAAFOVEKRLAV 159

QY 202 KYISIEPVVFIHIIITMTVEVLPVVTLCRDSAFSLGVTMLLT-CIWLKLVSAHTS 260
  - - - - -
Db 160 GALTEQAGLLHVNATILCFPAAVLLVESITPVGSLALMAHTILFLKLSYRDNVL 219

QY 261 DIR - - - - -SIANAADKANP - - - - -EVS - - - - -YVLSKSLAYFMVAPTLCYPSYPSACIRKG 309
  - - - - -
Db 220 WCRARAKAASAGKASSAAAPHTVSPDNLTIRDLTYFLFAPILCYELNPPSPRIRKR 279

QY 310 WYARQFAKLVITFGFMGFIIEQYINPIVRNSKHLKGLDLYA - - - - -IERVILKLVNLYVWL 367
  - - - - -
Db 280 FLRLRILEMLFTQVLQVLIQOMVPTIOMSKPKF-DMDYSRIERLLKLVNLYVWL 338

QY 368 CMFYCFHLNIIAELLCFGRDFYKDWNAKSVGDYWRWNNPVHKKWVRHIYFPCLR 427
  - - - - -
Db 339 IFYFWLHSCILNAELMQLFGDFEYRDNWNSSEVTYFWQNNIPVHKKWIRHEYFPMRLR 398

QY 428 SKIPKTLAIIITAFVSAVFHELCTAVPCRLKFLWAFGIMFQVPLV-FITVLOEFGST 486
  - - - - -
Db 399 RGSRRMARIGVFLASAFFHEYLVSPLRMFLWAFGMAQIPLANEVGRFFQGY - - - 455

QY 487 VGNMIFWIFCIGOPCMVLLYYHD 511
  - - - - -
Db 456 -GNAAVW-LTLIIGQPIAVLMYVHD 478

RESULT 3
DGT1_MOUSE
ID DGT1_MOUSE STANDARD; PRT; 498 AA.
AC Q9Z2A7; Q9D7Q5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Diacylglycerol O-acyltransferase 1 (EC 2.3.1.20) (Diglyceride

```

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 00:46:55 ; Search time 38 Seconds  
(without alignments)  
643.523 Million cell updates/sec

Title: US-09-623-514a-2  
Perfect score: 2771  
Sequence: 1 MAILDSAGVTTWNGGEF.....OPMCVLLYYHDLNMRKGSMS 520

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	791.5	28.6	488	1 DGT1_HUMAN	O75907 homo sapien
2	772.5	27.9	491	1 DGT1_CERAE	Q9gmf1 cercopithec
3	771.5	27.8	498	1 DGT1_MOUSE	Q9zza7 mus muscucu
4	762.5	27.5	498	1 DGT1_RAT	Q9erm3 rattus norv
5	415.5	15.0	540	1 SOA1_MOUSE	Q61263 mus muscucu
6	412.5	14.9	546	1 SOA1_CRIGR	Q60457 cricetus
7	406	14.7	550	1 SOA1_MACFA	O77761 macaca fasc
8	402	14.5	550	1 SOA1_CERAE	O77760 cercopithec
9	397	14.3	550	1 SOA1_HUMAN	P35610 homo sapien
10	391.5	14.1	545	1 SOA1_RAT	O70536 rattus norv
11	378.5	13.7	642	1 ARE2_YEAST	P53629 saccharomyc
12	375	13.5	525	1 SOA2_MOUSE	O88908 mus muscucu
13	357.5	12.9	522	1 SOA2_HUMAN	O75908 homo sapien
14	350.5	12.6	526	1 SOA2_CERAE	O77759 cercopithec
15	301	10.9	610	1 ARE1_YEAST	P23628 saccharomyc
16	277.5	10.0	537	1 AREH_SCHPO	Q10269 schizosacch
17	135	4.9	588	1 YAT1_SCHPO	O09758 schizosacch
18	129	4.7	560	1 YG14_YEAST	P53154 saccharomyc
19	122	4.4	715	1 NU5M_NEUCR	P05510 neurospora
20	120	4.3	579	1 KN2_HUMAN	Q9h2s1 homo sapien
21	120	4.3	1235	1 TRK1_YEAST	P21685 saccharomyc
22	115	4.2	652	1 NU5M_PODAN	P20679 podospira a
23	113	4.1	395	1 DLTB_BACSU	P39580 bacillus su
24	112	4.0	1241	1 TRK1_SACBA	P28569 saccharomyc
25	110	4.0	433	1 HXB3_MOUSE	P09026 mus muscucu
26	109.5	4.0	388	1 SR4_HUMAN	P31391 homo sapien
27	109	3.9	618	1 YKR4_YEAST	P36029 saccharomyc
28	107	3.9	633	1 Y147_HAEIN	P44543 haemophilus
29	106.5	3.8	574	1 KN2_MOUSE	P58390 mus muscucu
30	106	3.8	973	1 TRP5_HUMAN	Q9ul62 homo sapien
31	106	3.8	974	1 TRP5_RABIT	O62852 oryctolagus
32	106	3.8	975	1 TRP5_MOUSE	Q9qx29 mus muscucu
33	105.5	3.8	437	1 COAA_BPPFI	P25129 bacterioph

RESULT 1				
DGT1_HUMAN				
ID	DGT1_HUMAN	STANDARD;	PRT;	488 AA.
AC	O75907;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Diacylglycerol O-acyltransferase 1 (EC 2.3.1.20) (Diglyceride acyltransferase) (ACAT related gene product 1).			
DE	DGT1 OR DGAT OR AGRP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	MEDLINE=98434592; PubMed=9756920;			
RA	Oelkers P., Behari A., Cromley D., Billheimer J.T., Sturley S.L.;			
RT	"Characterization of two human genes encoding acyl coenzyme A:cholesterol acyltransferase-related enzymes.";			
RL	J. Biol. Chem. 273:26765-26771(1998).			
CC	-!- FUNCTION: Catalyzes the terminal and only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates.			
CC	-!- CATALYTIC ACTIVITY: Acyl-CoA + 1,2-diacylglycerol -> CoA + triacylglycerol.			
CC	-!- PATHWAY: Central role in the metabolism of cellular diacylglycerol lipids.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (By similarity).			
CC	-!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.			
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL; AF059202; AAC63997.1; --			
DR	Genew; HGNC:2843; DGAT1.			
DR	MIM; 604900; --			
DR	GO; GO:0008415; F:acyltransferase activity; TAS.			
DR	GO; GO:0004144; F:diacylglycerol O-acyltransferase activity; TAS.			
DR	GO; GO:0006641; P:triacylglycerol metabolism; TAS.			
DR	Pfam; PF03062; MBOAT; 1.			
KW	Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum.			
FT	TRANSMEM 104..124			
FT	TRANSMEM 130..150			
FT	TRANSMEM 166..186			
FT	TRANSMEM 189..209			
FT	TRANSMEM 282..302			
FT	TRANSMEM 332..352			
FT	TRANSMEM 406..426			
FT	TRANSMEM 428..448			

P70604 rattus norv  
P14651 homo sapien  
Q10254 schizosacch  
O9vxx3 drosophila  
O09029 mesocricetu  
P24881 ascaris suu  
P46094 homo sapien  
P46094 homo sapien  
P92953 arabidopsis  
P40351 saccharomyc  
Q9h255 homo sapien  
Q46378 chlamydia t

34 105.5 3.8 580 1 KN2\_RAT  
35 104.5 3.8 431 1 HXB3\_HUMAN  
36 104.5 3.8 445 1 YD26\_SCHPO  
37 103 3.7 694 1 FR22\_DROME  
38 102.5 3.7 367 1 BET3\_MESAU  
39 101 3.6 525 1 COX1\_ASCSU  
40 100.5 3.6 637 1 YHE7\_YEAST  
41 100 3.6 333 1 CXCL\_HUMAN  
42 99.5 3.6 318 1 ATH4\_ARATH  
43 99.5 3.6 577 1 ALG8\_YEAST  
44 98 3.5 320 1 OXE2\_HUMAN  
45 98 3.5 536 1 MVIN\_CHLTR

ALIGNMENTS

**THIS PAGE BLANK (USPTO)  
BEST AVAILABLE COPY**

**THIS PAGE BLANK (USPTO)**

DB 208 FYRENVPRKLNASKKSSVPIPTVNOYL-----YFELAPLILYKDSIPKPIPTVAMGTV 261  
 QY 312 ARQFAKLVITFGMGFIIEQYINPIVRNSK-HPKGDLLAIAERVLKLSVPMIYWLKMF 370  
 DB 262 AMQFAQVQGLFEEYVYIFERLCAPLEFRNKKQEPFSARVLV-----LCIF 305

## RESULT 14

565208  
 Probable membrane protein YPL189w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein p201

C:Species: Saccharomyces cerevisiae

C>Date: 10-Dec-1994 #sequence\_revision 31-May-1996 #text\_change 19-Apr-2002

C:Accession: S65208; S65201

R:Reger, M.; Mueller-Auer, S.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S65202

A:Accession: S65208

A:Molecule type: DNA

A:Residues: 1-609 <RIE>

A:Cross-references: EMBL:Z73545; NID:q1370394; PID:e246916; PID:q1370395; MIPS:YPL189w

A:Experimental source: strain S288C (AB972)

R:Benes, V.; Rechmann, S.; Nentwich, U.; Voss, H.; Ansoerge, W.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S65183

A:Accession: S65201

A:Molecule type: DNA

A:Residues: 177-609 <BEN>

A:Cross-references: EMBL:Z73545; MIPS:YPL189w

A:Experimental source: strain S288C (AB972)

C:Genetics:

A:Gene: SGD:GUP2

A:Cross-references: SGD:S0006110

A:Map position: 16L

C:Keywords: transmembrane protein

F:79-95/Domain: transmembrane #status predicted <TM1>

F:136-152/Domain: transmembrane #status predicted <TM2>

F:164-180/Domain: transmembrane #status predicted <TM3>

F:201-217/Domain: transmembrane #status predicted <TM4>

F:328-344/Domain: transmembrane #status predicted <TM5>

F:376-392/Domain: transmembrane #status predicted <TM6>

F:406-422/Domain: transmembrane #status predicted <TM7>

F:498-514/Domain: transmembrane #status predicted <TM8>

F:534-550/Domain: transmembrane #status predicted <TM9>

F:578-594/Domain: transmembrane #status predicted <TM10>

Query Match 5.8%; Score 163; DB 2; Length 609;  
 Best Local Similarity 19.9%; Pred. No. 3.5e-05;

Matches 94; Conservative 94; Mismatches 151; Indels 134; Gaps 27;

QY 123 SSSATFKQSHAGLFNLCVAVLLAVNSRLTIENLMKRYGMLIRDFWSSSLSDWPLFMQC 182  
 DB 122 NSDPQYRFRSNNFLLALILLOI---LTKRVFVFSKIPKRPDAGCL-----VFVGF 173  
 QY 163 I-----SLSTFPLA--AFVTEKLVLOKYSSEPVYIFLHITITWTEVLVYVYTLRCDASFL 236  
 DB 174 MGINSVKFLTHAFIEFTLHSLKRRRLIAFAFISYGIFTL-----FINQKKMLPF 226  
 QY 227 SGVTLMLLTCIYV-----LKLVSVA-----HTSYDIR-- 263  
 DB 227 NIIATILSPMDQWYGIYPRWDFPNFTLLRLILSTSMDFLERMHEQLSKPSIDYDDKRP 286  
 QY 264 ----SLANNA-----DK-----ANPEVSYYLSKSLAY-----FMVAPTL-- 294  
 DB 287 EFRKSLSGSTLOTIYESGKNVLEKEKRLVAEHIDYNNINIAITTYAPLVLGPIITF 346  
 QY 295 ---CYQPSYPSACIRKG---VVARQFAKLVITFGMGFIIEQYINPIVR---NSKHP 344  
 DB 347 NDYLQSENKLPISLTKKNIGFALYKVSLLIMEITLHYI---YGAIAIARTKAMNDPFL 403  
 QY 345 KQDLLAIAERVLKLSVPMIYWLKMFYCFPHMLNLA-----ELCCGDSREFYKDMWN 398

DB 404 Q-----QAMIALFNLI--NYLLILPWLFRIMAMVDGIDADENMLRCVDN-----N 449  
 QY 399 AASVDYWMNMNMPYKMWVRRIYEPCLSKIPKTLAIITAFVSAVHELCIAVPCRLF 458  
 DB 450 YSTVG-FRWAMHTSFNKWIRIYIVPFGSN-NKILTSFVSVFALIMHIDLRY---LF 504  
 QY 459 KLMAFLGIMFQVPLVFTN-YIQERPGSTVGMMTFEIF-CIFGQP--MCVLL 507  
 DB 505 --GWLTVLLILGELYITNCFSRNFRS-----WYRFVCGIGAINICMM 548

## RESULT 15

881409  
 Probable transmembrane transport protein Cj0611c [imported] - Campylobacter jejuni (s

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002

C:Accession: B81409

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kellley, J.M.; Churcher, C.; Basham, D.; Chli

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bar

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: B81409

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-458 <PAR>

A:Cross-references: GB:AL139075; GB:AL111168; NID:q6967817; PIDN:CAB75247.1; PID:q696

A:Experimental source: serotype O2, strain NCRC 11168

C:Genetics:

A:Gene: Cj0611c

Query Match 5.1%; Score 142; DB 2; Length 458;  
 Best Local Similarity 20.8%; Pred. No. 0.001;

Matches 86; Conservative 71; Mismatches 165; Indels 92; Gaps 20;

QY 136 ENLCVAVLLAVNSRLTIENLMKRYGMLIRDFWSSSLSDWPLFMCCISLIFPLAATV 195  
 DB 8 FSLIMIAFFAIY-----WTEKNDYRIQNLIL--IFSYIYILINPYFA--- 49  
 QY 196 EKLVILOKYSSEPVYIFLHITITWTEVLVYVYTLRCDASFL----- 236  
 DB 50 --LVLFYI-----TFPIHFALLIFVRRRIYFATC-MAFITLNLCFEYFPGSIVSDE 101  
 QY 237 ---SGVTLMLLTCIYVWLKLSVVAHTSYDIRSLANAADKANPEVSYYLSKSLAYFM-VA 291  
 DB 102 IINFGLEPLRIDVLPIDISFYTFSS--ITLVVEYQKRRE-----SFLMLAFLSRF 154  
 QY 292 PFLCYQPSYPSACIRKGVARQF--AKVIFTGPRGF-----IEQYINPIVRNSKHP 345  
 DB 155 PFLSGPIWRSSFFEQAOAKREKFRANLIIILVFGIVKVLNLYLGIYAKSIL--- 210  
 QY 346 GDLVLAIERVLKLSVPMIYWLKMFYCFPHMLNLI---AELLCGDSREFYDMNNAKSV 402  
 DB 211 -DPOSYNFIQLLSA--ITAVAIQIYCDSSGYDVLCAPALMGLFLLPNFMPIYAKNU 267  
 QY 403 GDYWRMNMNPPVHKMWVRRIYEPCLSKIPKTLAIITAFVSAVHELCIAVPCORLEK 459  
 DB 268 KDFMARWHLSLSTFIRDIYIPLGSRKGIPTVANIILAFILSGMWHGNTLA-----FI 322  
 QY 460 LMAFLGIMFQVPLVFTNLOERFG---SYGNNIMFWIFRIFQOPKCVLLXY 509  
 DB 323 VW---GLHIGIGIVFHLITLTSKFSLOKIPALGREFLTFQVGF---TWIFPY 369

Search completed: August 31, 2003, 04:59:56  
 Job time : 51 secs





## RESULT 9

T41684

Probable sterol O-acetyltransferase 2 - fission yeast (*Schizosaccharomyces pombe*)C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 01-Mar-2002

C:Accession: T41684

R:Medler, H.; Duesternhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1999

A:Reference number: 221742

A:Accession: T41684

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-472 &lt;MEM&gt;

A:Cross-references: EMBL:AL117183; PIDN:CA54864.1; GSPDB:GN00068; SPDB:SPCPIE11.05C

A:Experimental source: strain 972h; clone pl pIE11

C:Genetics:

A:Gene: SPDB:SPCPIE11.05C

A:Map position: 3

C:Superfamily: probable membrane protein YCR048w

Query Match

Best Local Similarity 11.9%; Score 329; DB 2; Length 472;

Matches 126; Conservative 74; Mismatches 194; Indels 76; Gaps 18;

QY 100 NADATF-----TTPSPVAHRRARESPLSDAITKOSH-----AGLENLGVVL 143  
 DB 15 NLEQTFKGVSETSKIDLRRAAYRPELSPTPS--IFARNYORNAVDFTGFVLEWAV 72  
 QY 144 IAVNSRLIIEMLMKGM-LIRDFWFSSRLSDMPFMCISLSIFPLAFLVEVLVQK 202  
 DB 73 SIMFMSLENFELTGRVYGTIRKYFQSNLID--LAAADLAMSMTFLAEPFOITPLG 130  
 QY 203 YIS-EPVYFLHITMTLVLYPYVTLRC--DSAFSGVTMLMLTCTVMLKLVSY-- 255  
 DB 131 YLWYGLGVLYXSLIL--LFLSHCVLRCLSNMSTHRAHMFILSHVILLKLSYVNV 187  
 QY 256 -AHNSYDRLS-----ANADKANPEVSY-----VLSKSLATYVAAPT 293  
 DB 188 NGWSTYCVHSLNKLQSKRTDLDDESSVEEYHCLNHGNTYPPENLITPALDELFMPS 247  
 QY 294 LCYQSPYRSACIR-----KGVVARQFAKLVIETFGMGIIEQYINPIYRNSKHPLK 345  
 DB 248 LCYQLYXPRTAHVRHRYIECALGFGCIFLLVITSDHFMVPLAKAIRITII--EAPD 304  
 QY 346 GDLLYALIER---VLKLSVPMLYWLMCFYCFHMLITLAEELCFGDRREYKDWMAKS 401  
 DB 305 ASATYFAIRLGHVAFLEMFPEMLSPFLVFWIEGVNCFSAITRFADRNFDWMNCWT 364  
 QY 402 VGDYWRMMNVHKKMNVHIFPCILRSKIPTLAIITLFLYSAYVHEL---CIAVPCLF 458  
 DB 365 WDOFARWYNNKPVHFLRHVYVP--LNSFMSKSLSTFFTFVSSVLAHELMGCITLKLINGY 423  
 QY 459 KLMAFLGIMFOVPLVETINYLOERFSGTVGNMIFETICIFGQPCVLLY 508  
 DB 424 GLFF---QMTQIPITIIOROKFVRNRHLGNIAIWFSL-IIGIALIALY 469

## RESULT 10

S19461

Probable membrane protein YCR048w - yeast (*Saccharomyces cerevisiae*)C:Species: *Saccharomyces cerevisiae*

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 20-Jun-2000

C:Accession: S19461; S19762

R:Grivell, L.A.; de Haan, M.; Maat, M.J.

submitted to the Protein Sequence Database, March 1992

A:Reference number: S19412

A:Accession: S19461

A:Molecule type: DNA

A:Residues: 1-610 &lt;GRI&gt;

A:Cross-references: EMBL:X59720; NID:G1907116; PIDN:CAA42296.1; PID:G1907190; MIPS:YCR04

R:Bohlin-Fukuhara, M.; Buhler, J.M.; Dajman-Fonfrier, B.; Dolira, C.; Francinques-Gallia

submitted to the Protein Sequence Database, March 1992

A:Reference number: S19396

A:Accession: S19762

A:Molecule type: DNA

A:Residues: 1-328 &lt;BOI&gt;

A:Cross-references: EMBL:X59720; MIPS:YCR048w

C:Genetics:

A:Gene: SGD:ARE1

A:Cross-references: SGD:S0000644; MIPS:YCR048w

A:Map position: 3R

C:Superfamily: probable membrane protein YCR048w

C:Keywords: transmembrane protein  
 F:186-202/Domain: transmembrane #status predicted <TM1>  
 F:228-255/Domain: transmembrane #status predicted <TM2>  
 F:262-284/Domain: transmembrane #status predicted <TM3>  
 F:289-305/Domain: transmembrane #status predicted <TM4>  
 F:370-387/Domain: transmembrane #status predicted <TM5>  
 F:401-425/Domain: transmembrane #status predicted <TM6>  
 F:453-469/Domain: transmembrane #status predicted <TM7>  
 F:538-555/Domain: transmembrane #status predicted <TM8>  
 F:590-606/Domain: transmembrane #status predicted <TM9>

Query Match

Best Local Similarity 10.9%; Score 301; DB 2; Length 610;

Matches 105; Conservative 76; Mismatches 161; Indels 82; Gaps 17;

QY 156 MKYGL-IR--TDFWSSRLSDMPFMCISLSIFPLAA--FTVERLYQKYSBPV 209  
 DB 195 MELGMAIRCTDYVAYSASA--WN-----KLEIVQYMTDPLFTIAMLDLAMELCFEPV 246  
 QY 210 IFLHIII-----TMTLVLYVYVTLKDCSAFLSGVTLMLTCTVW 249  
 DB 247 VFEHVLKRLIKMKMTGFVAVSIFELAFIPTEPIYV-YFDFENWVRIFLFLHSVVF 305  
 QY 250 LKIVSYA-----HTSYDIRSLANAADKANP-----EVSYY----- 279  
 DB 306 MKSHFAYNYGIMDKOLEYSSKQLOKYNESLSPERELIQRKSCDCELEINQOTDN 365  
 QY 280 -----VSLKSLAYFVAPTLCYQSPYRSACIRKGVVARQFAKLVIETFGMGIIEQY 333  
 DB 366 DFPNNISCSNPFMFLFVLVYQINYPRTSRIWRVYLEKVC-AITGITFLMAYTAQPFM 424  
 QY 334 NPI-----TRNSKHPKGLDLVAIERVLT---SVNLLV-NLCMYCFHMLNTLAEEL 385  
 DB 425 HPMARCIQHNTPFGGWIPTQGEWHLLDMIGFTVLVLTFTYMTDALLNCAVELT 484  
 QY 386 CFGDRREYKDWMAKSVDYWRMMNVHKKMNVHIFPCILRS-KIPTLAIITLFLYSA 444  
 DB 485 RFADRYFGDMWNCVSEFEFSKINNVPHKFLRHVHSSGALHLSQNTLFTFFLSA 544  
 QY 445 VFHELCTIAPCRLFLMAFLGIMFOVPLVETINYLOERFSGTVGNMIFETICIFGQPC 504  
 DB 545 VFHEKAMFAIRFRVAGYLFMFOLSOFTVATLSNTFLBARPOLSNVVSFGVC-SGPSII 603  
 QY 505 VLLY 508  
 DB 604 KTLV 607

## RESULT 11

T18744

Hypothetical protein B0395.2 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T18744

R:McMurry, A.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19014

A:Accession: T18744

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-467 &lt;MTL&gt;

A:Cross-references: EMBL:Z68131; PIDN:CAA92217.1; GSPDB:GN00028; CESP:B0395.2

A:Experimental source: clone B0395

## RESULT 7

A48026  
sterol O-acyltransferase (EC 2.3.1.26) - human  
N:Alternate names: ACAT; acyl-coenzyme A cholesterol acyltransferase  
C:Species: Homo sapiens (man)  
C>Date: 28-May-1999 #sequence, revision 28-May-1999 #text, change 05-May-2000  
C/Accession: A59038; A48026  
R:Chang, C.C.Y.; Chang, T.Y.  
submitted to Genbank, May 1999  
A:Description: Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol A:Accession: A59038  
A:Contents: correction  
A:Accession: A59038  
A:Molecule type: mRNA  
A:Residues: 1-550 <CHAA2>  
A:Cross-references: GB:I21934; NID:94878021; PIDN:ACG37532.2; PID:94878022  
R:Chang, C.C.Y.; Huh, H.Y.; Caddigan, K.M.; Chang, T.Y.  
J. Biol. Chem. 268, 20747-20755, 1993  
A:Title: Molecular cloning and functional expression of human acyl-coenzyme A:cholesterol A:Reference number: A48026; MUID:94012607; PMID:8407899  
A:Accession: A48026  
A:Molecule type: mRNA  
A:Residues: 1-206, 'R', 208-550 <CHAA1>  
A:Cross-references: GB:I21934  
C:Genetics:  
A:Gene: GDB:SOAT; STAT; ACAT  
A:Cross-references: GDB:251696; OMIM:102642  
A:Map position: 1q25-1q25  
C:Function:  
A:Description: catalyzes the esterification of cholesterol by acyl-CoA  
A:Pathway: cholesterol metabolism  
A>Note: helps maintain cellular cholesterol homeostasis; plays a role in the development  
C:Superfamily: sterol O-acyltransferase  
F:409/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Query Match

Best Local Similarity 14.3%; Score 397; DB 1; Length 550;  
Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

17 GGEFVLDLRKRRKSSDSNGL-----LLSGDNNSBDVGAADVDRIDSYVND 70  
70 GSHFDFVTLNLEKSSALDNGGALTFVLEGEKNNHRAKDRAPPE----- 117  
71 ACGTANAGDNNCGGNGGCGGCGGNADATFYRPSVPAHRRARRESPLSSDAIFK 130  
118 -GSKIRI-----ARSLIDELLVDHIRT 141  
131 SHAGFNLCAVVLIAV-----NSRLIE-NLKKYGM-LIRDFWESSRSLRDMPLMCC 182  
142 YHMFALLLLEFLSTLVVDYIDEGRLVLEFSLSTVAFGKPPYVW-----TWIMF--- 192  
183 ISLSTPLAFTVEKLVQKI-----SEPVYI-----FLHIIITMEVLY-PVYVTLR 230  
193 -----LSTFSPYVLEFQHWATGYKSSHPRLSLFHGLEFMIFQIGVGFPTVYL- 244  
231 CDSAPLGGVTLMLTICVILKIVSAHT---SYDIRSLANAADKAN-----PEVSYYSLK 283  
245 ---AYLLPPASRIILIFEDIRFMAKHSFVRENVPPVLSAKKSSVPIPIYNOQL--- 286  
284 SLAYENAVPTLCYOPSPSACIRKGVAAQFAKLVIFGFMGFIIEQYINPIVRSK-H 342  
299 ---YFLEAPTLTYRDSYPRNPVFMGVAMKFAQVGCFFYYVYIERICAPLFRNIKE 355  
343 PLKGDLLXIERLKLSV-----PNLYWLCMFYCFPHMLNLAELLCGDEFEYKDW 397  
356 PPSA-----RLVLICVENSILIPGLVLLFLEFFAFHCLMNAFAMLRGDMFYM 408  
398 NAKSVGVYRMNMPVHKMMVNRHLY--PCLRSKIKRKTALIIIAFLVSAVHELCIAVQC 455  
409 NSTSYSNYRTVMVYVHDMLYIAYADFLMFPSKRKSAAMLAVFASVAVHEIYALAV-C 467

## RESULT 8

456 R-----LFLKMAFLGIMQVPLVFTITNLYQERFGSTVGMMIRW-----FICFIQG 501  
468 LSFEPVPLVFLVFMFPGMAFN-----FLVNDKRK---PIWVLMWMTSLFLGNGVLICFYSQ 520

## Query Match

Best Local Similarity 13.7%; Score 378.5; DB 2; Length 642;  
Matches 133; Conservative 83; Mismatches 176; Indels 135; Gaps 19;

102 DATEFYRPSV-----PAHRRARE-----SPLSSDAIF 128  
125 DMFEHRPSLSDSVNEPFKTRFVGPPLKEIRRREKELAMAKNLHHRSSPDAYDSVG 184  
129 KQSHAGFLNLCV-----VLIAVNSRLIEN-----LMTYGM-----IRTFWFS-SRS 172  
185 KNDGAAPTYPTATSETVYVETTLIISNFSGLYAFWMAIFAGVAKALIDYIOHNGS 244  
173 LRDMPLMCCISLSTPLAFTVEKLVQKVISEPVYIFH-----IIT 217  
245 FKDSER-----LKFMTNLFTVAVDLMNTLSTYFVGVIOYCKMGVLKMGTTGWTFS 298  
218 MFEVLYPVVYTLRCDNA-----PLSGVTLMLTICVILKIVSA-----HNSYDI 262  
299 IYEFLEFVIFVYLTENILKLMHLSKIFLPHSLVLLMKHMSFAYNGYLMGIEELQFSK 358  
263 RSLANAADKAN-PEV-----SIYSLSLAVFMAVPTL 294  
359 SALAKYKDSINDKRVIGALEKSECFSELSOSLSDOQKFPNNISAKSFETFTPTL 418  
295 CYOPSPSACIRKGVAAQFAKLVIFGFMGFIIEQYI-NPI-----VRNSKHPLKGD 347  
419 IYIEYPRKTEIKWSVLEKIC-AIRGTTIYLMADQILMTVPAMRALAVRNS-----E 471  
348 LLYAIEIRVKT-----SVPNLYV-NLCMFYCFPHMLNLAELLCGDEFEYKDW 400  
472 WNGIILRLKMWGLADVDYGPVIMVILIDBYLLMDALINCVALTFRGDRYFGDMWNCV 531  
401 SVGDYRMNMPVHKMMVNRHLYPCRS-KIPTALIIIAFLVSAVHELCIAVPCRLK 459  
532 SWADESRINIWIPIVHKEFLRHVYHSSMSSEFLNKSQATLMTFFLSVYHELAIVYIKRL 591  
460 IMAFLGIMQVPLVFTITNLYQERFGSTVGMMIRWFCIFGQPMCVL 506  
592 FYLEFFQMLQPLVALTITKFMNRRTIIGVIMWLDICMGPSVWCTL 638

Best Local Similarity 31.5%; Pred. No. 1,6e-50;  
Matches 166; Conservative 98; Mismatches 166; Indels 97; Gaps 14;

QY 27 RRRKSRSDSNGLLSGSDNNSPSDVGAPADVRIDSVNDAGTANLAGDNNGGD 86  
DB 7 RRRKQPSERTSNGSLASSR-----SSFAQ 30  
QY 87 NNGGGRGEGCGGNADATFTYPSVPAHRRARESPSSDAIRKQSHA-----GLEMLCVV 141  
DB 31 NNNSSRKSEEMRGPEKV-----VH-----TAQDSLSTSGMTNFRGFNLSTIL 75  
QY 142 VLIANSRLITENLKKYGLITFDWFSS-----RSLRWPLFMCCISISTIFPLAFYVE 196  
DB 76 LVLVLSNGRALENKIVIGLITPLQWISTEVEHHYSIMWPMIALILCSNIDILSFGKE 135  
QY 197 KLVLOKTYSEPVVILHIIITTEVLVYVYTLRCD-SAFLSGVTLMLTICVIMLKVSY 255  
DB 136 KILRGWGLONGFAAFYISLVIAHLIPVYVTLTKHKKRPLMSVVMGYVIEALKETISY 195  
QY 256 AHTSY---DIR-----SLANAADKA-----NPEVSY-----VSLKSLA 286  
DB 196 GHANVWARDARRKITELKTQVTDLAKKTCDFKQFMDLKDLSMHQMAAQYPAHFLTSNIT 255  
QY 287 YVMAVPTLCYQSPYSRACIRKGVAROPAKIVITGFGFTIEQYINIVANSKHPKLG 346  
DB 256 YVMAVPTLCYQSPYSRACIRKGVAROPAKIVITGFGFTIEQYINIVANSKHPKLG 314  
QY 347 DLLVA--IERVYKLSVPLNYVMICMFCFHLMLNLIALCFGEDEFFKDMWNAKSVGD 404  
DB 315 EMEYRCLERLKLAIAPNLHILVLFYTFHFSLMLIALDLFADREFRDMNAETTCY 374  
QY 405 YVRMNMVPHKVMVNHITPCLRSKIPKTLAIIIFLVSAYFHELICIAVPCRLFKLMAFL 464  
DB 375 FFKSNMIVPHRAVHVISPMARNNFSKSAFFVVEFVSAPFHEVYSVPLKIFRLMSY 434  
QY 445 GTFQVPLVETINYLQERGSVGMIFEFICIFGQPCVLLYHD 511  
DB 435 GMMGQIPLSTITDKVVR--GGRGNITVW--LSLIVGQPLAILMYGD 478

## RESULT 5

T01293  
hypothetical protein F27F23.25 - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 30-Apr-1999  
C:Accession: T01293  
R:Roundsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
submitted to the EMBL Data Library, May 1998  
A:Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.  
A:Reference number: Z14177  
A:Accession: T01293  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-131 <ROU>  
A:Cross-references: EMBL:AC003058; NID:g3135250; PID:g3135275  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Map position: 2  
A:Note: F27F23.25

Query Match 24.2%; Score 670; DB 2; Length 131;  
Best Local Similarity 100.0%; Pred. No. 8.6e-47;  
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGYTVTEGGGEFVDLRLRRKRSRSDSNGLLSGSDNNSPSDVGAPADVR 60  
DB 1 MAILDSAGYTVTEGGGEFVDLRLRRKRSRSDSNGLLSGSDNNSPSDVGAPADVR 60  
QY 61 DRIDSVNDAGTANLAGDNNGGDNNGGGEGGGRGNADATFTYPSVPAHRRARS 120  
DB 61 DRIDSVNDAGTANLAGDNNGGDNNGGGEGGGRGNADATFTYPSVPAHRRARS 120  
QY 121 PLSSDAIRKO 130

DB 121 PLSSDAIRKO 130

## RESULT 6

sterol O-acetyltransferase (EC 2.3.1.26) - mouse  
N:Alternate names: acyl-coenzyme A cholesterol acyltransferase  
C:Species: Mus musculus (house mouse)  
C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 02-Jun-2000  
C:Accession: I49454; J04617  
R:Reijnen, P.J.; Oka, K.; Sullivan, M.C.; Chang, T.  
J. Biol. Chem. 270, 26192-26201, 1995  
A:Title: Molecular cloning of mouse ACAT.  
A:Reference number: I49454; M0ID:96064687; PMID:7592824  
A:Accession: I49454  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-540 <RES>  
A:Cross-references: GB:I42293; NID:g1066809; PIDN:MAC42075.1; PID:g1066810  
R:Green, S.; Steinberg, D.; Quehenberger, O.  
Biochem. Biophys. Res. Commun. 218, 924-929, 1996  
A:Title: Cloning and expression in Xenopus oocytes of a mouse homologue of the human  
A:Reference number: J04617; M0ID:96158386; PMID:8579615  
A:Accession: J04617  
A:Molecule type: mRNA  
A:Residues: 1-194/R', 196-540 <GRE>  
A:Cross-references: GB:S81092; NID:g1478335; PIDN:AB36050.1; PID:g1478336  
C:Comment: This enzyme helps maintain cellular cholesterol homeostasis by catalyzing  
y cholesterol and oxygenated sterols.  
C:Genetics:  
A:Gene: ACAT  
A:Map position: 1  
C:Superfamily: sterol O-acetyltransferase  
C:Keywords: acyltransferase; cholesterol; coenzyme A; endoplasmic reticulum; transmem

Query Match 15.0%; Score 415.5; DB 1; Length 540;  
Best Local Similarity 28.5%; Pred. No. 1.3e-25;  
Matches 135; Conservative 76; Mismatches 149; Indels 113; Gaps 24;

QY 99 GNADATFTYRSPVPAHRA-----RESPSSDAIRKQSHA----- 134  
DB 81 GCAITFTSILEMKKNNRKKDLRAPEQKIFISROSL--DELFEDVDIRIYHMFIAL 138  
QY 135 --LFNLCVAVLAVN--SRLIE-NLMKYGM--LIRDFWSSSLRDMPLFMCCISLSTPP 189  
DB 139 LILFVLSTIVVDYIDBGRVLVEPNLILAFGRFPVYV-----TWAMFSLTSLTPV- 191  
QY 190 LAAFVVERLVLOKYL-----SEPVVIFL-HIITMTEVL-----YPVVTLRCDAPLS 237  
DB 192 -----LFQPMHAGYSKSSHPLISLVHGLFLVPLDGLVGFPTVVL---AYTL 238  
QY 238 GVTMLILCTIVWLKLYVAHT--SYDIRLANAADRAN-----PEVSYVSLKSLAYENV 290  
DB 239 PPASRFLILDEQIRILMKRSHFVRREIPVLANAEKSSKDLPLPVNOVL-----YFLF 292  
QY 291 APTLCYQSPYSRACIRKGVAROPAKIVITGFGFTIEQYINIVANSKHPKLG-HPLKGDLL 349  
DB 293 APTLIRYDNYFPTPYRMCYVAMQFLQVGCIFYIYIFERLCADIPRNIKQEPISA--- 349  
QY 350 VALIERVLKLSV-----PNLYVMICMFCFHLMLNLIALCFGEDEFFKDMWNAKSVGD 404  
DB 350 -----RVLVLCVFNSTLPGVLLIFLSEFAFLHCLMNAFAMFLGDBMFGKDMWNSTSYSN 405  
QY 405 YVRMNMVPHKVMVNHITPCL--RSKIPKTLAIIIFLVSAYFHELICIAVPCR----- 456  
DB 406 YRTNWNVVDWLVVYVVKDLMLMFFSKRKRKSAAMLAVALFASVVEHETALAI-CLSYFVY 464  
QY 457 LFKLNAFLGIMQVPLVETINYLQERGSVGMIFW-----FIFCIQGO 501  
DB 465 LFLVLFMFGMAFN-----FIVNDSKR--PIMNIMWASLPLGGLILCFYSQ 510

QY 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYIQ 480  
 |||||  
 Db 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYIQ 480  
 |||||  
 QY 481 ERGSGTVGNMIFEFICIFGQPCVLLYYHDLNRRGSM 520  
 |||||  
 Db 481 ERGSGTVGNMIFEFICIFGQPCVLLYYHDLNRRGSM 520  
 |||||

## RESULT 2

H84576

diacylglycerol O-acyltransferase [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C&gt;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001

C/Accession: H84576

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shua, T.P.; Bentto, M.T.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A&gt;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: H84576

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 &lt;STO&gt;

A:Cross-references: GB:AE002093; NID:94191775; PIDN:AA01044.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g19450

A:Map position: 2

Query Match 81.4%; Score 2255.5; DB 2; Length 441;  
 Best Local Similarity 83.5%; Pred. No. 1.1e-173;  
 Matches 434; Conservative 5; Mismatches 2; Indels 79; Gaps 3;

QY 1 MAIIDSAGTYTENGSGFVDLRLRRRSRSDSSNGILSGSDNPSDDVGAPADVR 60  
 |||||  
 Db 1 MAIIDSAGTYTENGSGFVDLRLRRRSRSDSSNGILSGSDNPSDDVGAPADVR 60  
 |||||  
 QY 61 DRIDSVDNDAGCTANLADGNNGGGNGGRCGRCGNADATFTYRSPARRARRES 120  
 |||||  
 Db 61 DRIDSVDNDAGCTANLADGNNGGGNGGRCGRCGNADATFTYRSPARRARRES 120  
 |||||  
 QY 121 PLSSDAIFRQSHAGFNLCCVVLIAVNSRLIENLMKYGLITDFWSSRLRDPLEW 180  
 |||||  
 Db 121 PLSSDAIFRQSHAGFNLCCVVLIAVNSRLIENLMKYGLITDFWSSRLRDPLEW 180  
 |||||  
 QY 181 CCISLSIFPLAFTVEKVLQKISEPVVIFLIIITMTVEVLPYVTLRCDSAFLSGVT 240  
 |||||  
 Db 181 CCISLSIFPLAFTVEKVLQKISEPVVIFLIIITMTVEVLPYVTLRCDSAFLSGVT 240  
 |||||  
 QY 241 LMLTCTIWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFWAPLICYOPSY 300  
 |||||  
 Db 241 LMLTCTIWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFWAPLICYOPSY 300  
 |||||  
 QY 179 LMLTCTIWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFWAPLICYOPSY 238  
 |||||  
 Db 179 LMLTCTIWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFWAPLICYOPSY 238  
 |||||  
 QY 301 PPSACIRKQWVARQFAKLVITFGFNGFTIEQYINPIVRNSKHPKLDLLAIERVILKSY 360  
 |||||  
 Db 301 PPSACIRKQWVARQFAKLVITFGFNGFTIEQYINPIVRNSKHPKLDLLAIERVILKSY 360  
 |||||  
 QY 239 PPSACIRKQWVARQFAKLVITFGFNGFTIEQYINPIVRNSKHPKLDLLAIERVILKSY 298  
 |||||  
 Db 239 PPSACIRKQWVARQFAKLVITFGFNGFTIEQYINPIVRNSKHPKLDLLAIERVILKSY 298  
 |||||  
 QY 361 PNLTYWLCMEYCFEHLMLNIIAELLCFGRREFKDMWNKSVGDYRMMNMPVHKMMVHR 420  
 |||||  
 Db 361 PNLTYWLCMEYCFEHLMLNIIAELLCFGRREFKDMWNKSVGDYRMMNMPVHKMMVHR 420  
 |||||  
 QY 299 PNLTYWLCMEYCFEHLMLNIIAELLCFGRREFKDMWNKSVGDYRMMNMPVHKMMVHR 358  
 |||||  
 Db 299 PNLTYWLCMEYCFEHLMLNIIAELLCFGRREFKDMWNKSVGDYRMMNMPVHKMMVHR 358  
 |||||  
 QY 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYIQ 480  
 |||||  
 Db 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYIQ 480  
 |||||  
 QY 481 ERGSGTVGNMIFEFICIFGQPCVLLYYHDLNRRGSM 520  
 |||||  
 Db 481 ERGSGTVGNMIFEFICIFGQPCVLLYYHDLNRRGSM 520  
 |||||

## RESULT 3

T01294  
 sterol O-acyltransferase homolog F27F23.26 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 30-Apr-1999  
 C/Accession: T01294

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K  
 submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.

A:Reference number: Z14177

A:Accession: T01294

A>Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-231 <R0U>

A:Cross-references: EMBL:AC003058; NID:93135250; PID:93135276

A:Experimental source: cultivar Columbia

C:Genetics:

A:Map position: 2

A:Insertions: 44/3; 70/3; 104/3; 150/2; 177/3; 184/3; 205/3

A>Note: F27F23.26

Query Match 43.4%; Score 1202; DB 2; Length 231;  
 Best Local Similarity 97.4%; Pred. No. 2.7e-89;  
 Matches 222; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 230 RCDSAFLSGVTMLLFCIWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFW 289  
 |||||  
 Db 3 RCDSAFLSGVTMLLFCIWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFW 62  
 |||||  
 QY 290 VAPLICYOPSPRSACIRKQWVARQFAKLVITFGFNGFTIEQYINPIVRNSKHPKLDLL 349  
 |||||  
 Db 63 VAPLICYOPSPRSACIRKQWVARQFAKLVITFGFNGFTIEQYINPIVRNSKHPKLDLL 122  
 |||||  
 QY 350 YAIERVLKLSVPMIYWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFW 409  
 |||||  
 Db 123 YAIERVLKLSVPMIYWLKIVSAHTSYDIRSLANADKANPEVSYVSLKSLAYFW 182  
 |||||  
 QY 410 NMPVHKMMVHRHIFPCRSKIPKTLAIIIAFLVSAVHELCI--AVPC 455  
 |||||  
 Db 183 NMPVHKMMVHRHIFPCRSKIPKTLAIIIAFLVSAVHELCI--AVPC 230  
 |||||

## RESULT 4

T19027

probable sterol O-acyltransferase (EC 2.3.1.26) H19N07.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 05-May-2000

C/Accession: T19027; T23106

R:Kershaw, J.

submitted to the EMBL Data Library, June 1996

A:Accession: T19027

A:Reference number: Z19061

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-496 <WIL>

A:Cross-references: EMBL:Z19061; PIDN:CAA99773.1; GSPDB:GN00023; CESP:H19N07.4

A:Experimental source: clone C06H2

R:Dobson, R.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z19678

A:Accession: T23106

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-496 <W12>

A:Cross-references: EMBL:Z29835; PIDN:CAB07399.1; GSPDB:GN00023; CESP:H19N07.4

A:Experimental source: clone H19N07

C:Genetics:

A:Gene: CESP:H19N07.4

A:Map position: 5

A:Insertions: 5/3; 43/2; 77/3; 139/1; 171/3; 456/1

C:Superfamily: sterol O-acyltransferase

C:Keywords: acyltransferase; coenzyme A

Query Match 26.1%; Score 724.5; DB 2; Length 496;

P. N. S. S. S.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: August 31, 2003, 04:34:04 ; Search time 46 Seconds  
(without alignments)  
1087.124 Million cell updates/sec

Title: US-09-623-514a-2  
Perfect score: 2771  
Sequence: 1 MAILDSAGVTVTENGGEF.....QPMCVLLYHDLNRRKGSMS 520

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues  
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_76:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	2	diacylglycerol O-a
2	2255.5	81.4	441	2	diacylglycerol O-a
3	1202	43.4	231	2	sterol O-acyltrams
4	724.5	26.1	496	2	probable sterol O-
5	670	24.2	131	2	hypothetical prote
6	415.5	15.0	540	1	sterol O-acyltrams
7	397	14.3	550	1	sterol O-acyltrams
8	378.5	13.7	472	2	probable membrane
9	329	11.9	472	2	probable sterol o-
10	301	10.9	610	2	probable membrane
11	279	10.1	467	2	hypothetical prote
12	277.5	10.0	537	2	probable sterol O-
13	165.5	6.0	305	2	sterol O-acyltrams
14	162	5.8	609	2	probable membrane
15	142	5.1	458	2	probable membrane
16	139.5	5.0	404	2	probable membrane
17	134.5	4.9	865	2	probable membrane
18	129	4.7	560	2	probable membrane
19	126	4.5	495	2	probable membrane
20	122	4.4	715	2	probable membrane
21	121.5	4.4	450	2	conserved hypothet
22	121.5	4.4	450	2	conserved hypothet
23	121.5	4.4	1170	2	probable 2-acylly
24	120	4.3	1235	1	potassium transpor
25	118.5	4.3	865	2	probable membrane
26	117.5	4.2	499	2	alinate o-acetyl
27	116	4.2	1689	2	sodium channel pro
28	115	4.2	652	2	NADH2 dehydrogenas
29	113	4.1	395	2	diLB protein - Bac

30	112	4.0	352	2	T48903	wax synthase (limp
31	112	4.0	1241	2	JU0466	potassium transpor
32	111	4.0	395	2	B96610	hypothetical prote
33	111	4.0	420	2	DB3556	probable coat prot
34	109.5	4.0	263	2	AE1557	hypothetical prote
35	109.5	4.0	388	2	JN0605	somatostatin recep
36	109	3.9	618	1	S38004	probable transport
37	108.5	3.9	783	2	A96825	hypothetical prote
38	108	3.9	527	2	G64626	alinate O-acetyla
39	107	3.9	334	2	T19955	hypothetical prote
40	107	3.9	395	2	AE2349	hypothetical prote
41	107	3.9	633	2	T64143	hypothetical prote
42	106.5	3.8	1553	2	T18502	hypothetical prote
43	106	3.8	546	2	S52053	cytochrome-c oxida
44	106	3.8	2895	2	H85362	hypothetical prote
45	105.5	3.8	437	2	S15144	hypothetical prote

ALIGNMENTS

Used this for R (ist)  
as reference  
500

RESULT 1	
T52584	diacylglycerol O-acyltransferase (EC 2.3.1.20) [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)	
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 17-Nov-2000	
C:Accession: T52584	
R:Hills, M.J.; Lu, C.; Hobbs, D.H.	
FEBS Lett. 452, 145-149, 1999	
A:Title: Cloning of a cDNA encoding diacylglycerol acyltransferase from Arabidopsis t	
A:Reference number: 226127	
A:Accession: T52584	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-520 <Hills>	
A:Cross-References: EMBL:AJ131831; PIDN:CA644774.1	
A:Experimental source: cultivar Columbia	
C:Function:	
A:Description: EC 2.3.1.20 [validated, MUID:99313150]; catalyzes the final acylation	
A:Pathway: triacylglycerol biosynthesis	
C:Keywords: acyltransferase; coenzyme A	
Query Match	100.0%; Score 2771; DB 2; Length 520;
Best Local Similarity	100.0%; Pred. No. 4.5e-215;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 MAILDSAGVTVTENGGEFVDLRLRRKSSDSSNGLLSGSNNSPSDVGAADVR 60
DB	1 MAILDSAGVTVTENGGEFVDLRLRRKSSDSSNGLLSGSNNSPSDVGAADVR 60
OY	61 DRIDSVNDADAGTANLADGNNGGDNNGGGEGRGADATFTYRSPVAHRRARS 120
DB	61 DRIDSVNDADAGTANLADGNNGGDNNGGGEGRGADATFTYRSPVAHRRARS 120
OY	121 PLSDAIFKOSHAGLFNLCVVLIVNSRLITENIMKRYGWLIRTFWESSRSLRMPLE 180
DB	121 PLSDAIFKOSHAGLFNLCVVLIVNSRLITENIMKRYGWLIRTFWESSRSLRMPLE 180
OY	181 CCISSISIFPLAFYVEKVLQKYSIEPVYIEFHIIITMEVLYPYVYLIRCSAFLSGVT 240
DB	181 CCISSISIFPLAFYVEKVLQKYSIEPVYIEFHIIITMEVLYPYVYLIRCSAFLSGVT 240
OY	241 LMLFCIYWLKIVSYAHTSYDRLSANAADKRNPEVSYSLSLAFVAVATPLCYOPSY 300
DB	241 LMLFCIYWLKIVSYAHTSYDRLSANAADKRNPEVSYSLSLAFVAVATPLCYOPSY 300
OY	301 PSACIRKGVARQFAKIVFTGFMGFIIEQYINPIVRNSKHPGLDLYAERVLKLSV 360
DB	301 PSACIRKGVARQFAKIVFTGFMGFIIEQYINPIVRNSKHPGLDLYAERVLKLSV 360
OY	361 PRLIYWLKMFYCFEFLIMLILAEELCFDGRREYKRWMAKSGVDYWRMNMVHHWVRH 420
DB	361 PRLIYWLKMFYCFEFLIMLILAEELCFDGRREYKRWMAKSGVDYWRMNMVHHWVRH 420

**THIS PAGE BLANK (USPTO)**

RESULT 15				
CENB_HUMAN				
ID	CENB_HUMAN	STANDARD;	PRT;	599 AA.
AC	P07199;			
DT	01-APR-1988 (Rel. 07, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Major centromere autoantigen B (Centromere protein B) (CENP-B).			
CN	CENPB.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=91372020; PubMed=1893793;			
RT	"CENP-B is a highly conserved mammalian centromere protein with			
RL	homology to the helix-loop-helix family of proteins."			
RN	Chromosome 100:360-370(1991).			
RP	[2]			
RX	SEQUENCE FROM N.A.			
RX	MEDLINE=21638749; PubMed=11780052;			
RA	DeJouas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Stavrides G., Almeida J.P., Babbage A.K., Begguley C.L.,			
RA	Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA	Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,			
RA	Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graffham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,			
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,			
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,			
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,			
RA	Lehvesaaho M.H., Levestra M.A., Lloyd C., Lloyd D.M., Lovell J.D.,			
RA	Marsh V.L., Martin S.L., McConachie L.J., McLay K., McMuray A.A.,			
RA	Malne S.A., Mistry D., Moore M.J.F., Mullikin J.C., McKernan T.,			
RA	Olliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,			
RA	Plummer B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,			
RA	Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,			
RA	Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,			
RA	Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,			
RA	Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,			
RA	Whitehead S.L., Whiteaker P., Willey D.L., Williams L., Williams S.A.,			
RA	Wilmink L., Wray P.W., Hubbard T., Dublin R.M., Bentley D.R., Beck S.,			
RA	Rogers J.,			
RT	The DNA sequence and comparative analysis of human chromosome 20."			
RL	Nature 414:865-871(2001).			
RN	[3]			
RP	SEQUENCE OF 6-599 FROM N.A.			
RX	MEDLINE=87166180; PubMed=2435739;			
RA	Barnshaw W.C., Sullivan K.F., Nachlin P.S., Cooke C.A.,			
RA	Kaiser D.A., Pollard T.D., Rothfield N.F., Cleveland D.W.,			
RT	"Molecular cloning of cDNA for CENP-B, the major human centromere			
RL	autoantigen."			
RL	J. Cell Biol. 104:817-829(1997).			
RN	[4]			
RP	SUBUNITS AND DOMAINS.			
RX	MEDLINE=93107144; PubMed=1469042;			
RA	Yoda K., Kitagawa K., Masumoto H., Muro Y., Okazaki T.,			
RT	"A human centromere protein, CENP-B, has a DNA binding domain			
RT	cont			

```

RT "A helix-turn-helix structure unit in human centromere protein B
(CENP-B)." ;
RL EMBL J_17:827-837(1998).
CC -I- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN
CC CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE
CC DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE
CC SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS
CC CENTROMERE FORMATION AND KINETOCHORE ASSEMBLY IN MAMMALIAN
CC CHROMOSOMES.
CC -I- SUBUNIT: Homodimer.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- SIMILARITY: Contains 1 CENPB domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X55039; CAA36879.1; -.
DR EMBL; AL109804; CAC17547.1; -.
DR EMBL; X05299; CAA28918.1; -.
DR PIR; S18735; S18735.
DR PDB; 1BM6; 07-OCT-98.
DR PDB; 1HIV; 11-JAN-02.
DR Genew; HGNC:1852; CENPB.
DR MIR; 117140; -.
DR GO; GO:0005698; C:centromere; NAS.
DR GO; GO:0003682; F:chromatin binding activity; NAS.
DR GO; GO:0003696; F:satellite DNA binding activity; NAS.
DR GO; GO:0000669; F:centromere/kinetochore complex maturation; NAS.
DR InterPro; IPR004875; CENP-B.
DR InterPro; IPR006600; CENPB.
DR Pfam; PF04218; CENP-B_N; 1.
DR Pfam; PF03184; DDE; 1.
DR SMART; SM00674; CENPB; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Centromere;
KW 3D-structure.
FT DNA_BIND 1 125 GLU-RICH (ACIDIC).
FT DOMAIN 404 465 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 508 538 R->M (IN REF. 3).
FT CONFLICT 583 583 R->M (IN REF. 3).
FT CONFLICT 592 593 VR->LL (IN REF. 3).
FT HELIX 10 21
FT HELIX 28 35
FT TURN 39 40
FT TURN 41 47
FT HELIX 41 47
FT HELIX 49 52
FT HELIX 53 53
FT TURN 53 53
SQ SEQUENCE 599 AA; 65171 MW; 9B4B7DB957A914AA CRC64;

Query Match 16.0%; Score 138; DB 1; Length 599;
Best Local Similarity 26.3%; Pred. No. 0.11;
Matches 36; Conservative 18; Mismatches 37; Indels 46; Gaps 3;

QY 38 EEEQQQLROGEEHTARNGSVVVEVRPGQNDSDGOGLFEENNRRFISVDSS-GNQE 96
Db 406 EEEEEEEEEEBEGEGBEGEEEGBEGEEBEGGEELGEEBE-----VEEGDVDSDE 458
QY 97 EOEEDERHAGE-----OOEDEEEHEEM 118
Db 459 EEEEDDESSSGLEAEADMAQGVVEAGGSFGAYGAQEAQCPTLFLRGEDSDSDSEED 518
QY 119 DEESDDPDQSDDSSRED 135
Db 519 DEEDEDDEDDDDDEED 535

```



SQ SEQUENCE 765 AA; 89509 MW; 006A20A7F968DB6A CRC64;  
 Query Match 16.2%; Score 139.5; DB 1; Length 765;  
 Best Local Similarity 22.8%; Pred. No. 0.11;  
 Matches 41; Conservative 38; Mismatches 58; Indels 43; Gaps 6;  
 QY 1 MN0ELLVSGSRRRTGSLRGNPSSQVDEQNMRYVEEEOQ-----QLR 46  
 DB 593 LKEMWVIGSRWRI-----SOSQKHYKTLAEBOQRQYKHLIDLWVKSLSFQDR 642  
 QY 47 QOEHEHTARN---GEVGVGPRPGQNDGQQLLENNNRFIISVDEDSGNOEEOEDE 102  
 DB 643 AAYEYISNKRKNMTKLRGPKRSRTTLQSKSESEDDDE---EEDDEEEEEEEDD 698  
 QY 103 EH-----AGQDEDEDEE---EEMDQESDDPQSDSRDEDEHTNVTSSSLVD 151  
 DB 699 ENGDSSEDDGDDSSSESESEDEDDDDDDDEDDDEDDDEDDNESESSSSSSSGD 758  
 RESULT 14  
 ID CNM4 BOVIN STANDARD; PRT; 1394 AA.  
 AC Q28181; Q03861; Q28082;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 240 kDa protein of rod photoreceptor CNM-channel [Contains: Glutamic acid-rich protein (GARP); Cyclic nucleotide-gated cation channel 4 (CNM channel 4) (CNM-4) (Cyclic nucleotide-gated cation channel modulatory subunit)].  
 DE CNM1 OR CNM4.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.  
 NC NCB1\_TaxID=9913;  
 [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=96009859; PubMed=7546742;  
 RA Goerz S., Colville C., Mueller F., Dose A., Godde M., Molday L., Kaupp U.B., Molday R.S.;  
 RA "A 240 kDa protein represents the complete beta subunit of the cyclic nucleotide-gated cation channel from rod photoreceptor";  
 RT Neuron 15:627-636 (1995).  
 RL [2]  
 RP SEQUENCE OF 454-1394 FROM N.A.  
 RC TISSUE=Testis;  
 RX MEDLINE=96198098; PubMed=8626431;  
 RA Biel M., Zong X., Ludwig A., Sauter A., Hofmann F.;  
 RA "Molecular cloning and expression of the modulatory subunit of the cyclic nucleotide-gated cation channel";  
 RT J. Biol. Chem. 271:6349-6355 (1996).  
 RL [3]  
 RP SEQUENCE OF 1-590 FROM N.A.  
 RC TISSUE=Retina;  
 RA Sugimoto Y., Yatsunami K., Tsujimoto M., Khorana H.G., Ichikawa A.;  
 RA Submitted (XX-1991) to the EMBL/GenBank/DBJ databases.  
 CC -1 SUBUNIT: FORMS FUNCTIONAL HETEROLOGIC CHANNELS WITH CNM3.  
 CC -1 SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1 ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Comment=Isoform CNM4D is the most frequent isoform  
 CC (CNM4D:CNM4C:CNM4E = 20:2:1) in testis;  
 CC Name=CNM4C;  
 CC IsoId=Q28181-1; Sequence=Displayed;  
 CC Name=CNM4D;  
 CC IsoId=Q28181-2; Sequence=VSP\_001109;  
 CC Name=CNM4E;  
 CC IsoId=Q28181-3; Sequence=VSP\_001108;  
 CC -1 TISSUE SPECIFICITY: RETINA, TESTIS, KIDNEY, HEART, AND BRAIN.  
 CC -1 SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X89626; CA61769.1; -  
 CC EMBL; X94707; CA64367.1; -  
 CC EMBL; M61185; AAA30536.1; -  
 CC PIR; A40437; A40437.  
 DR InterPro: IPR000595; CNM\_binding.  
 DR InterPro: IPR005821; Ion\_trans.  
 DR Pfam; PF00027; CNM\_binding; 1.  
 DR Pfam; PF00520; Ion\_trans; 1.  
 DR SMART; SM00100; CNM; 1.  
 DR PROSITE; PS00888; CNM\_BINDING\_1; 1.  
 DR PROSITE; PS00889; CNM\_BINDING\_2; 1.  
 DR PROSITE; PS50042; CNM\_BINDING\_3; 1.  
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane; Multigene family; Alternative splicing.  
 KW CHAIN 1 590  
 FT CHAIN 454 1394  
 FT DOMAIN 1 767  
 FT TRANSMEM 768 786  
 FT DOMAIN 801 819  
 FT TRANSMEM 820 844  
 FT DOMAIN 845 901  
 FT TRANSMEM 902 924  
 FT DOMAIN 925 968  
 FT TRANSMEM 969 988  
 FT DOMAIN 1073 1093  
 FT TRANSMEM 1094 1394  
 FT NP\_BIND 1081 1219  
 FT BINDING 1141 1141  
 FT CARBOHYD 1067 1067  
 FT VAAPPLIC 515 532  
 FT VAAPPLIC 522 530  
 FT CONFLICT 341 341  
 FT CONFLICT 454 465  
 FT CONFLICT 482 482  
 FT CONFLICT 499 499  
 FT CONFLICT 572 590  
 FT CONFLICT 1283 1283  
 FT CONFLICT 1289 1289  
 FT CONFLICT 1336 1336  
 FT CONFLICT 1338 1338  
 SQ SEQUENCE 1394 AA; 155064 MW; BE6DA559BE374AA7 CRC64;  
 Query Match 16.1%; Score 138.5; DB 1; Length 1394;  
 Best Local Similarity 24.6%; Pred. No. 0.22;  
 Matches 42; Conservative 35; Mismatches 43; Indels 51; Gaps 7;  
 QY 17 GSLRGNPSSQV--DEQNMRYVE-----EEQOQLRQOEHTARNGEV 59  
 DB 321 GSTSTSPRTSAPADEKGVVEQTPELPRIQEKDEBEKEDGEEBEKGEKKEE 380  
 QY 60 VG--VEPPGQNDGQQLLENNNRFIISV-----EDSGNOEEOE---DERHAGEQ 108  
 DB 381 EGEKEGEKGEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKEKKE 440  
 QY 109 DEED-----EEMDQESD-----DPOSDDSSDE 136  
 DB 441 EEEGRGKEVEGREDEBEQDHSVLDSYVPSQSEEDRSSESTODQ 491

```

RA Loebl A-M., Seaver L., Bonnefont J.-P., Romano C., Ficheta M.,
RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.,
RT "Evaluation of a mutation screening strategy for sporadic cases of
RT ATR-X syndrome.";
RL J. Med. Genet. 36:183-186 (1999).
RN (17)
RN VARIANTS ATR-X SER-179; LEU-190; ILE-194; CYS-246; PHE-1552; SER-1645
RP AND CYS-1847.
RX MEDLINE=20451413; PubMed=10995512;
RA Wada T., Kubota T., Fukushima Y., Satoh S.;
RT "Molecular genetic study of Japanese patients with X-linked alpha-
RT thalassemia/mental retardation syndrome (ATR-X).";
RL Am. J. Med. Genet. 94:242-248 (2000).
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN BRAIN
CC DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC HEREROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC INTERACTING WITH HPL.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=1;
CC IsoId=P46100-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P46100-2; Sequence=VSP_000575;
CC Name=3;
CC IsoId=P46100-3; Sequence=VSP_000574;
CC Name=4;
CC IsoId=P46100-4; Sequence=VSP_000576;
CC Name=5;
CC IsoId=P46100-5; Sequence=VSP_000574, VSP_000576;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DISEASE: Defects in ATRX are the cause of X-linked alpha-
CC thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-
CC X is an X-linked disorder comprising severe psychomotor
CC retardation, facial dysmorphism, urogenital abnormalities, and
CC alpha-thalassemia. An essential phenotypic trait are hemoglobin H
CC erythrocyte inclusions.
CC -1- DISEASE: Defects in ATRX are the cause of Sutherland-Haas X-linked
CC mental retardation syndrome (SHS) [MIM:309470]. It is
CC characterized by severe mental retardation with spastic
CC paraplegia, microcephaly, short stature and cryptorchidism.
CC -1- DISEASE: Defects in ATRX are a cause of Smith-Pinman-Myers
CC syndrome (SPM) [MIM:309580]. Clinical features include severe
CC mental retardation, microcephaly, growth failure, facial anomalies
CC and bilateral cryptorchidism. Due to the clinical overlap with
CC ATR-X syndrome, some patients originally diagnosed as having SPM,
CC might be affected by a variant of ATR-X syndrome which lack
CC hemoglobin h inclusions.
CC -1- DISEASE: Defects in ATRX are the cause of Carpenter-Waziri
CC syndrome (CWS), an X-linked recessive condition characterized by
CC moderate mental retardation, short stature, brachydactyly with
CC excessive skin creases, and widening of the knuckles.
CC -1- DISEASE: Defects in ATRX are the cause of Juberger-Marsidi syndrome
CC (JM) [MIM:309590]. JM is a rare X-linked recessive disease
CC characterized by severe mental retardation, growth failure,
CC sensorineural deafness, microgenitalism and early death.
CC -1- SIMILARITY: BELONGS TO THE SNR2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
-----
Query Match 16.5%; Score 141.5; DB 1; Length 2492;
Best Local Similarity 27.4%; Pred. No.0.26;
Matches 49; Conservative 21; Mismatches 42; Indels 67; Gaps 7
-----
12 RRRTGGSLRGNPSSQVDEQGNRV-----VEEEOOQOLROQ-----EE 50
Db 1300 KKRFGKQVNEBPG-----DEAKNQVNSSDSDSESKPRRHRRLRLKLTYSDDSEGE 135
51 EHT--ANRGEVGVVYEPKRGQNDQSO-----GQOLEEN 80

```

```

Db      | 1356 KTTPEKHKEVKNRNKRVSSDESDDFQSGCVSESVSESDERQRTPSAKAELEEN 1415
Oy      | 81 -----NNRFISVDSSGNOEBOEDDEHAGCODEDEEHMDOESDDPQSDS 131
Db      | 1416 QRSYKKKKRRIRIKVQDSSSENKSNGSEEBEEBEKEEBEEEEEBEEBEE-----DENDS 1470

RESULT 13
UBF1_MOUSE STANDARD; PRT; 765 AA.
ID UBF1_MOUSE
AC P25976;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1).
GN UBF1 OR TCFUBF OR UBF1 OR UBF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (UBF1 AND UBF2).
RC STRAIN=C3H/He;
RX MEDLINE=91367658; PubMed=1891354;
RA Hiatsake K., Nishimura T., Maeda Y., Hanada K.I., Song C.Z.,
   Muramatsu M.;
RT "Cloning and structural analysis of cDNA and the gene for mouse
RT transcription factor UBF.";
RL Nucleic Acids Res. 19:4631-4637(1991).
CC -I- FUNCTION: UBF RECOGNIZES THE RIBOSOMAL RNA GENE PROMOTER AND
CC ACTIVATES TRANSCRIPTION MEDIATED BY RNA POLYMERASE I THROUGH
CC COOPERATIVE INTERACTIONS WITH THE SPECIES-SPECIFIC FACTOR SL1. IT
CC BINDS SPECIFICALLY TO THE UPSTREAM CONTROL ELEMENT.
CC -I- SUBUNIT: Homodimer.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=UBF1; Synonyms=Long;
CC IsoId=P25976-1; Sequence=Displayed;
CC Name=UBF2; Synonyms=Short;
CC IsoId=P25976-2; Sequence=VSP_002194;
CC -I- SIMILARITY: Contains 6 HMG box domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60831; CAA43222.1; -.
DR PIR; S22314; S22314.
DR HSBP; P07155; HMF.
DR MGd; MG1:98512; Ubtf.
DR GO; GO:0005730; C:nucleolus; IDA.
DR GO; GO:0005515; F:protein binding activity; IPI.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 5.
DR SMART; SMO0398; HMG; 6.
DR PROSITE; PSS0118; HMG_BOX_2; 6.
KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Repeat; Alternative splicing.
FT     DNA_BIND    112       180      HMG BOX 1.
FT           196       264      HMG BOX 2.
FT     DNA_BIND    298       362      HMG BOX 3.
FT     DNA_BIND    407       475      HMG BOX 4.
FT     DNA_BIND    482       549      HMG BOX 5.
FT     DNA_BIND    568       634      HMG BOX 6.
FT     DOMAIN      675       765      ASF/GLU/SBR-RICH (ACIDIC).
FT     VARSPLIC    221       257      Missing (in isoform UBF2).
FT                                     /Fld-VSP_002194.
```

DB 643 AAYKEYISNKKRNMKLRGNPKSGRTTLQSKSESEDDDD-----EEDDDDDDEEEEDDD 697  
 QY 103 EH-----AGDEDEDEDEEEMDQSDPDQSD--SSNDEHTHTNSTNNS 148  
 DB 698 ENGDSSEDEGDSSESSSESEDESDGDNEDDDDDDEDEDEDEDEDESESSSS 753

RESULT 12  
 ATRX\_HUMAN  
 ID ATRX\_HUMAN STANDARD; PRT; 2492 AA.  
 AC P46100; P51068; Q15886; Q9H021; Q9NTS3;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked  
 nuclear protein) (XNP) (Zn-HX).  
 GN ATRX OR RAD54L OR XH2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 NCBI\_taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1, 2, 3, 4, AND 5), VARIANT SER-1860, AND  
 RP VARIANTS ATR-X.  
 RX MEDLINE=97123494; PubMed=8968741;  
 RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,  
 RA Gibbons R.J.;  
 RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations  
 RT point to a common mechanism underlying the ATR-X syndrome.";  
 RL Hum. Mol. Genet. 5:1899-1907(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).  
 RX MEDLINE=97386582; PubMed=9244431;  
 RA Villard L., Lissi A.-M., Cardoso C., Proud V., Chiaroni P.,  
 RA Colleaux L., Schwartz C., Fontes M.;  
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding  
 RT a potential zinc finger helicase.";  
 RL Genomics 43:149-155(1997).  
 RN [3]  
 RP SEQUENCE OF 860-2492 FROM N.A.  
 RX MEDLINE=95179111; PubMed=7874112;  
 RA Stayton C.L., Dabovic B., Giuliano M., Gecz J., Broccoli V.,  
 RA Giovannazzi S., Bossoiasco M., Monaco L., Raetan S., Boncinelli E.,  
 RA Bianchi M.E., Consalez G.G.;  
 RT "Cloning and characterization of a new human Xq13 gene, encoding a  
 RT putative helicase.";  
 RL Hum. Mol. Genet. 3:1957-1964(1994).  
 RN [4]  
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=94214473; PubMed=8162050; Villard L., Stayton C.L.,  
 RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,  
 RA Millaesau P., Khrestchatsky M., Fontes M.;  
 RT "Cloning and expression of the murine homologue of a putative human  
 RT X-linked nuclear protein gene closely linked to PKX1 in Xq13.3.";  
 RL Hum. Mol. Genet. 3:39-44(1994).  
 RN [5]  
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.  
 RX MEDLINE=95211835; PubMed=7697714;  
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;  
 RT "Mutations in a putative global transcriptional regulator cause X-  
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";  
 RL Cell 80:837-845(1995).  
 RN [6]  
 RP SEQUENCE OF 1375-2492 FROM N.A.  
 RX Pearce A., Chapman J.;  
 RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP EZH2 BINDING.  
 RX MEDLINE=98167853; PubMed=9499421;  
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,  
 RA Colleaux L.;  
 RT "Specific interaction between the XNP/ATR-X gene product and the SET  
 RT domain of the human EZH2 protein.";

RL Hum. Mol. Genet. 7:679-684(1998).  
 RN [8]  
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC  
 RP HETEROCHROMATIN.  
 RX MEDLINE=20040663; PubMed=10570185;  
 RA McEwell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,  
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,  
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;  
 RT "Localization of a putative transcriptional regulator (ATRX) at  
 RT pericentromeric heterochromatin and the short arms of acrocentric  
 RT chromosomes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).  
 RN [9]  
 RP DISEASE.  
 RX MEDLINE=20233147; PubMed=10751095;  
 RA Villard L., Fontes M., Ades L.C., Gecz J.;  
 RT "Identification of a mutation in the XNP/ATR-X gene in a family  
 RT reported as Smith-Fineman-Myers syndrome.";  
 RL Am. J. Med. Genet. 91:83-85(2000).  
 RN [10]  
 RP VARIANT ATR-X SER-1713.  
 RX MEDLINE=97196774; PubMed=9043863;  
 RA Villard L., Lacombe D., Fontes M.;  
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype  
 RT without alpha-thalassemia.";  
 RL Eur. J. Hum. Genet. 4:316-320(1996).  
 RN [11]  
 RP VARIANT JM GIN-2131.  
 RX MEDLINE=96224392; PubMed=8630485;  
 RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,  
 RA Munnich A., Lyonnet S.;  
 RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";  
 RL Nat. Genet. 12:359-360(1996).  
 RN [12]  
 RP VARIANTS ATR-X.  
 RX MEDLINE=97467722; PubMed=9326931;  
 RA Gibbons R.J., Bachoo S., Picketts D.J., Affimos S., Asenbauer B.,  
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,  
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Stanley S.F.,  
 RA Higgs D.R.;  
 RT "Mutations in transcriptional regulator ATRX establish the functional  
 RT significance of a PHD-like domain.";  
 RL Nat. Genet. 17:146-148(1997).  
 RN [13]  
 RP VARIANT ATR-X LEU-246.  
 RX MEDLINE=20123062; PubMed=10660327;  
 RA Fichera M., Romano C., Castiglita L., Falla P., Ruberto C., Amata S.,  
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;  
 RT "New mutations in XNP/ATR-X gene: a further contribution to  
 RT genotype/phenotype relationship in ATR/X syndrome.";  
 RL Hum. Mutat. 12:214-214(1998).  
 RN [14]  
 RP VARIANT SHS-LYS-1742.  
 RX MEDLINE=99347960; PubMed=10417298;  
 RA Lissi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,  
 RA Prieto F., Fontes M., Martinez F.;  
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental  
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:  
 RT bias.";  
 RL Am. J. Hum. Genet. 65:558-562(1999).  
 RN [15]  
 RP VARIANT CWS THR-2050.  
 RX MEDLINE=99326061; PubMed=10398237;  
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,  
 RA Curtis M.;  
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";  
 RL Am. J. Med. Genet. 85:249-251(1999).  
 RN [16]  
 RP VARIANTS ATR-X GLU-175, 178-VAL--LYS-198 DEL; SER-190; PRO-219;  
 RP LEU-246 AND CYS-249.  
 RX MEDLINE=99219535; PubMed=10204841;  
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougué J.,

```

Db      253 YEQDERDQKEGNDYDTRSEADSGSESISFTD 285

RESULT 10
ID      MP62_LYTP1      STANDARD;      PRT;      411 AA.
AC      P91753;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Mitotic apparatus protein p62.
OS      Lytechinus pictus (Painted sea urchin).
OC      Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC      Echinoidea; Echinoidea; Echinacea; Temnopneustidae;
OC      Lytechinus;
NCBI_taxid=7653;
(1)
SEQUENCE FROM N.A.
RP      MEDLINE=97166213; PubMed=9013612;
RA      Ye X., Sloboda R.D.;
RT      "Molecular characterization of p62, a mitotic apparatus protein
RT      required for mitotic progression."
RL      J. Biol. Chem. 272:3606-3614(1997).
CC      -1- FUNCTION: REQUIRED FOR MITOTIC PROGRESSION. BINDS TO CHROMATIN.
CC      -1- SUBCELLULAR LOCATION: Nuclear
CC      -1- PTM: PHOSPHORYLATED BY CAM-KINASE II IN VITRO.
CC      -1- SIMILARITY: BELONGS TO THE NUCLEOPHOSMIN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
CC      EMBL; U76750; AAB47481.1; -
DR      InterPro; IPR004301; Nucleoplasmn.
KW      P1am; PF03066; Nucleoplasmn; 1.
KW      Nuclear protein; Phosphorylation; Mitosis.
FT      DOMAIN 147 156 POLY-GLU.
FT      DOMAIN 192 200 POLY-GLU.
FT      DOMAIN 262 267 POLY-GLU.
FT      DOMAIN 272 284 POLY-GLU.
FT      DOMAIN 293 300 POLY-ASP.
FT      DOMAIN 313 319 POLY-GLU.
SQ      SEQUENCE 411 AA; 46385 MW; D2DB04293C362254 CRC64;

Query Match      16.8%; Score 144.5; DB 1; Length 411;
Best Local Similarity 25.5%; Pred. No. 0.031;
Matches 42; Conservative 33; Mismatches 45; Indels 45; Gaps 8;

QY      3 QELTVSGKRR-----RTGSLNGNPSSQVDEQNMRRVVEEQQQLRQOEENHTRN 56
DB      155 EETPKGSGPKKIVKIAVKGKRMKG--KGDELDU-----EDDEEEEEEELIQTAKG 206
QY      57 GEVGVGEPRPGQNDSDQ-----GQLENNNRFTSVDESSGNQ----- 95
DB      207 KK-----RAPSAKAGKAKTLAVVDGTSKRKVPNGSV--NGHALDDDEDEBDDYKVG 260
QY      96 EEQEDDEHAGEQDEDEDEEEMQESD---DPQDSDSRDE 136
DB      261 DEEBEEATSGEEEBEEDDEEEDDEEMALGDDDDDEDEDEDE 305

RESULT 11
UBFI_RAT      STANDARD;      PRT;      764 AA.
AC      P25977; P25978;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)

```

```

DE Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1).
DN UBF1 OR TCBT8 OR UBF-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus,
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (UBF1 AND UBF2).
RA MEDLINE=91195316; PubMed=2014238;
RA O'Mahony D.J., Rothblum L.I.,
RT "Identification of two forms of the RNA polymerase I transcription
RT factor UBF."
RN Proc. Natl. Acad. Sci. U.S.A. 88:3180-3184(1991).
RN [2]
RP SUBUNIT.
RA MEDLINE=99199559; PubMed=10099786;
RA Riddsdale R.A., Semcock J.L., Larson D.E., Rothblum L.I., Haraux G.;
RT "Topology of recombinant rat upstream binding factor."
RN Biochem. Cell Biol. 76:649-655(1998).
CC -1- FUNCTION: UBF RECOGNIZES THE RIBOSOMAL RNA GENE PROMOTER AND
CC ACTIVATES TRANSCRIPTION MEDIATED BY RNA POLYMERASE I THROUGH
CC COOPERATIVE INTERACTIONS WITH THE SPECIES-SPECIFIC FACTOR SL1. IT
CC BINDS SPECIFICALLY TO THE UPSTREAM CONTROL ELEMENT.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=UBF1; Synonyms=Long;
CC IsoId=P25977-1; Sequence=Displayed;
CC Name=UBF2; Synonyms=Short;
CC IsoId=P25977-2; Sequence=VSP_002195;
CC -1- SIMILARITY: Contains 6 HMG box domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, M61726; -, NOT_ANNOTATED_CDS.
CC DR EMBL, M61725; -, NOT_ANNOTATED_CDS.
CC DR PIR, A40439; A40439.
CC DR PIR, B40439; B40439.
CC DR HSSP, P07155; IHMF.
CC InterPro: IPR000910; HMG_box_5.
CC Pfam: PF00505; HMG_box_5.
CC SMART, SM00398; HMG_6.
CC DR PROSITE, PS50118; HMG_BOX_2; 6.
CC Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Repeat; Alternative splicing.
FT DNAStrand 112 180 HMG_BOX_1.
FT DNAStrand 196 264 HMG_BOX_2.
FT DNAStrand 298 362 HMG_BOX_3.
FT DNAStrand 407 475 HMG_BOX_4.
FT DNAStrand 482 549 HMG_BOX_5.
FT DNAStrand 568 634 HMG_BOX_6.
FT DNAStrand 675 764 ASP/GLU/SER-RICH (ACTIDIC).
FT DNAStrand 765 825 ASP/GLU/SER-RICH (ACTIDIC).
FT VARSPLIC 221 257 Missing (in isoform UBF2).
FT FTId=VSP_002195.
FT SEQUENCE 764 AA; 89437 MW; B9D3J71615DB0534 CRC64;
Query Match 16.5%; Score 141.5; DB 1; Length 764;
Batch local Similarity 23.9%; Pred. No. 0.084;
Matches 42; Conservative 35; Mismatches 56; Indels 43; Gaps 6;
OY 1 MNOELLVSGSKRRRTGSLRGNPSSQVDEQNRNVVEEQOQ-----QLR 46
D5 : : : : : : : : : : : : : : : : : : : : : : : : : : :
593 LKEMVVEIGSMQGI-----SSQSKENHYKKLAEFGQGRQKKYHLDLVKYSLSPODR 642
OY : : : : : : : : : : : : : : : : : : : : : : : : : : :
47 QQEEHTARN---GEVGVVEPRPGGNDSSQGGLENNNNRFTSVDESSGNGEEDEDE 102
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```

[illegible]

DT 15-SEP-2003 (Rel. 42, Last annotation update)  
DE Putative splicing factor YT521 (RA301-binding protein) .

GN YTS21.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxId=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Brain;  
RX MEDLINE=9814187; PubMed=9473574;  
RA Imai Y., Matsuo N., Osawa S., Tohyama M., Takagi T.;  
RT "Cloning of a gene, YT521, for a novel RNA splicing-related protein  
induced by hypoxia/reoxygenation."  
RL Brain Res. Mol. Brain Res. 53:33-40(1998).

RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RX MEDLINE=20032183; PubMed=10564280;  
RA Hartmann A.M., Nayler O., Schwarzer F.W., Obermeyer A., Stamm S.;  
RT "The interaction and colocalization of Sam68 with the splicing-  
associated factor YT521-B in nuclear dots is regulated by the Src  
family kinase p59Lyn."  
RL Mol. Biol. Cell 10:3909-3926(1999).  
CC -1- FUNCTION: May be part of a signal transduction pathway that  
influences splice site selection.  
CC -1- SUBUNIT: Interacts with SPR510/RA301, SC35, SR2 and Sam68.  
CC -1- SUBCELLULAR LOCATION: Nuclear; nucleoplasm.  
CC -1- ALTERNATIVE PRODUCTS:  
Event=Alternative splicing; Named isoforms=2;  
Name=1; Synonyms=YT521-B;  
IsoId=G9YX02-1; Sequence=Displayed;  
Name=2;  
IsoId=G9YX02-2; Sequence=VSP\_006819, VSP\_006820;

CC CC -1- TISSUE SPECIFICITY: Ubiquitous.  
CC CC -1- PTM: Tyrosine-phosphorylated.  
CC CC -1- SIMILARITY: Contains 1 YTH domain.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See http://www.ebi.ac.uk/announce/  
or send an email to license@ebi.ac.uk).

-----  
DR EMBL; D78303; BAA23885.1; --  
DR EMBL; AF044731; AAC55973.1; --  
DR Pfam; PF04146; YTH: 1.  
DR PROSITE; PSS0882; YTH: 1.  
KW Nuclear protein; mRNA splicing; Alternative splicing; Phosphorylation.  
FT DOMAIN 358 495 YTH.  
FT FT DOMAIN 172 263 GLU-RICH.  
FT FT DOMAIN 612 654 PRO-RICH.  
FT FT DOMAIN 658 738 ARG-RICH.  
FT FT VARSPPLIC 328 345 Missing (in isoform 2).  
FT FT VARSPPLIC 577 584 /FtId=VSP\_006819.  
FT FT MISSING (in isoform 2).  
SQ SEQUENCE 738 AA; 85825 MW; 220D0CDBA22A7FD CRC64;

Query Match 16.9%; Score 145; DB 1; Length 738;  
Best Local Similarity 28.8%; Fred. No. 0.051;  
Matches 44; Conservative 25; Mismatches 66; Indels 18; Gaps 4

DQ 14 RTGSLRNPSSGYQDECMRVVEEEOOQLROEEHTARNGEVGVPRPGQNQSQ 73  
Db KRKDSERAKAPTPGPSRRIGLEVDNRASRSSGSKEE---GNSERYSDHETSSASSE 192  
Dq OGOLRENN-----NRFISVEDBSSGNQ-----EQEDEEHAGODEEDDEEMD 119  
Db QGNNTNEEGGEDEVEDEEDGDDEDEADAEDEEDBEDEDEDEEEEEEEE 252  
DQ QESDDPFGSDDSREDETH-TNSTVNSSIIVD 151



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AL355013; CAB90146.1; -  
 DR EMBL; Z98762; CAB11485.1; -  
 DR EMBL; Z98599; CAB11250.2; -  
 DR EMBL; D89140; BAA13802.1; -  
 DR PIR; T38786; T38786;  
 DR PIR; T42417; T42417;  
 DR GeneDB; SPOMbe; SPAC4A8.16c; -  
 DR InterPro; IPR000717; PCI.  
 DR SMART; SM00088; PINT; 1.  
 KW Initiation factor; Protein biosynthesis.  
 FT DOMAIN 1 188 ASP/GLU/SER-RICH.  
 FT CONFLICT 683 683 M -> I (IN REF. 2).  
 SQ SEQUENCE 918 AA; 104367 MW; 43B70599409688BF CRC64;

Query Match 17.4%; Score 149.5; DB 1; Length 918;  
 Best Local Similarity 24.0%; Pred. No. 0.034;  
 Matches 36; Conservative 36; Mismatches 65; Indels 13; Gaps 4;

QY 16 GGSILGNPSSQVDEQNNRVV-----EEQQOQLRQEEHTARNGEVGVPEPPGQGN 70  
 DB 7 GGS--SPSDASVVSSEENRLTSSRLKKQDSSSESESESSASSSESESE-----ES 60  
 QY 71 DSQOGLFENNRRFISVDESSGQGEDEDEHAGDEDEDEDEDEDEDEDEDDSDDFDSD 130  
 DB 61 ESESEVEVPKKKVAASEDESE 118

QY 131 SSRDEHTNTSVNSSIVDLPVHQLSSP 160  
 DB 119 ESE 148

RESULT 6  
 UBF1\_HUMAN STANDARD; PRT; 764 AA.  
 AC P17480.  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-AUG-1990 (Rel. 15, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1)  
 DE (Anticentriolar NOR-90)  
 GN UBF1 OR UBF1 OR UBF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RX MEDLINE=90231434; PubMed=2330041;  
 RA Jantzen H.M., Admon A., Bell S.P., Tjian R.;  
 RT "Nucleolar transcription factor hUBF contains a DNA-binding motif  
 RT with homology to HMG proteins";  
 RL Nature 344:830-836(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A. (UBF1).  
 RX MEDLINE=92044316; PubMed=1940801;  
 RA Chan E.K.L., Imai H., Hamel J.C., Tan E.M.;  
 RT "Human autoantibody to RNA polymerase I transcription factor hUBF.  
 RT Molecular identity of nucleolus organizer region autoantigen NOR-90  
 RT and ribosomal RNA transcription upstream binding factor";  
 RL J. Exp. Med. 174:1235-1244(1991).  
 CC -1- FUNCTION: RECOGNIZES THE RIBOSOMAL RNA GENE PROMOTER AND ACTIVATES  
 CC TRANSCRIPTION MEDIATED BY RNA POLYMERASE I THROUGH COOPERATIVE  
 CC INTERACTIONS WITH THE SPECIES-SPECIFIC FACTOR SL1. IT BINDS  
 CC SPECIFICALLY TO THE UPSTREAM CONTROL ELEMENT.  
 CC -1- SUBUNIT: Homodimer.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- ALTERNATIVE PRODUCTS:  
 CC Event-Alternative splicing; Named isoforms=2;  
 CC Name=UBF1; Synonyms=Long;  
 CC IsoId=PI17480-1; Sequence=Displayed;  
 CC Name=UBF2; Synonyms=Short;  
 CC IsoId=PI17480-2; Sequence=VSP 002193;  
 CC -1- SIMILARITY: Contains 6 HMG box domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X53390; CAA37469.1; -  
 DR EMBL; X53461; CAA37548.1; -  
 DR EMBL; X56687; CAA40016.1; -  
 DR PIR; S09318; S09318.  
 DR PIR; S18193; S18193.  
 DR PDB; 1K99; 04-DEC-02.  
 DR TRANSFAC; T02900; -  
 DR Genew; HGNC:12511; UBF1.  
 DR MIM; 600673; -  
 DR GO; GO:0005730; C:nucleolus; TAS.  
 DR GO; GO:0003701; F:RNA polymerase I transcription factor activity; TAS.  
 DR GO; GO:0006356; P:regulation of transcription from Pol I prom. ; TAS.  
 DR InterPro; IPR000910; HMG\_12\_box.  
 DR Pfam; PF00505; HMG\_box; 5.  
 DR SMART; SM00398; HMG; 6.  
 DR PROSITE; PSS0118; HMG\_BOX\_2; 6.  
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein;  
 KW Repeat; Alternative splicing; 3D-structure.

FT DNA\_BIND 112 180 HMG\_BOX 1.  
 FT DNA\_BIND 196 264 HMG\_BOX 2.  
 FT DNA\_BIND 298 362 HMG\_BOX 3.  
 FT DNA\_BIND 407 475 HMG\_BOX 4.  
 FT DNA\_BIND 482 549 HMG\_BOX 5.  
 FT DNA\_BIND 568 634 HMG\_BOX 6.  
 FT DOMAIN 675 764 ASP/GLU/SER-RICH (ACIDIC).  
 FT VARSPIC 221 257 Missing (in isoform UBF2).  
 FT /Ftrd-VSP 002193.  
 FT SEQUENCE 764 AA; 89406 MW; D4F08BE180E757D CRC64;

Query Match 17.1%; Score 147; DB 1; Length 764;  
 Best Local Similarity 23.2%; Pred. No. 0.04;  
 Matches 41; Conservative 35; Mismatches 57; Indels 44; Gaps 5;

QY 1 MNDELIVSGSKRRRTGSLGNPSSQVDEQNNRVVEEQOQ-----QLR 46  
 DB 593 LKEMVETIGSRWRI-----SSQKEHYKLAEEQOKYKHLDMVSLSPDR 642

QY 47 QQEEHTARNGEVV-----GVEPPGQNDSCQGLENNRRFISVDESSGQGEDEDE 102  
 DB 643 AAYEYISNKRKSMTKLRGNPKSSRTTLOSKSESEDDDE-----DDDDDEDEDEDEDE 698

QY 103 -----EHAGEDEDEDEDEDEDEDDFDDSDSRDEHTNTSVN 147  
 DB 699 NGDSESDGDSSESSSE 755

RESULT 7  
 MT1\_HUMAN STANDARD; PRT; 1121 AA.  
 ID MT1\_HUMAN Q01538; Q94922; Q9UPV2;  
 AC 01-JUN-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Myelin transcription factor 1 (MT1) (MT1) (Proteolipid protein  
 DE binding protein) (PLPB1).  
 GN MT1 OR PLPB1 OR MT1 OR MT1 OR KIAA0835 OR KIAA1050.  
 OS Homo sapiens (Human).

QY 3 QELLVSGSKRRRTG-----GSLRGNPSSQVDEQMNVRVEEQQOQLRQEEHTAR 55  
 DB 65 EEEVEEEERREERREGEEREGEGREAEEREEAEEREEAEEREEAE 124  
 QY 56 NGEVVGEPGPGGNDQOQOLENNRFTSVDESSGNGOEDEDEHAGEOEDEEE 115  
 DB 125 EAE---AEEERAEEREEAEEREEAEEREEAEEREEAEEREEAE 181  
 QY 116 EEMDQESDDPQDSDSSREDE 136  
 DB 182 EEEAEAEAEAEAEAEAEAEAE 202

RESULT 4  
 CENB\_CRIGR STANDARD; PRT; 606 AA.  
 ID CENB\_CRIGR  
 AC P48988;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Major centromere autoantigen B (Centromere protein B) (CENP-B).  
 GN CENP-B.  
 OS Cricetus griseus (Chinese hamster).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Cricetulus.  
 NC NCBI\_TaxID=10029;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=96254058; PubMed=8652663;  
 RA Bejarno L.A., Valdivia M.M.;  
 RT "Molecular cloning of an intronless gene for the hamster centromere antigen CENP-B";  
 RL Biochim. Biophys. Acta 1307:21-25 (1996).  
 CC -1- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS CENTROMERE FORMATION AND KINETOCHORE ASSEMBLY IN MAMMALIAN CHROMOSOMES (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Contains 1 CENP domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/submit/> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL: U20951; AAB06494.1; -.  
 CC PIR: S70358; S70358.  
 CC DR HSSP: P07199; 1BM6.  
 CC DR InterPro: IPR004875; CENP-B.  
 CC DR InterPro: IPR006600; CENP-B.  
 CC DR Pfam: PF04218; CENP-B\_N; 1.  
 CC DR Pfam: PF03184; DDE; 1.  
 CC DR SMART: SM00674; CENP-B; 1.  
 CC KW Chromosomal protein; Nuclear protein; DNA-binding; Centromere.  
 CC FT DNA BIND 1 125 BY SIMILARITY.  
 CC FT DOMAIN 404 471 GLU-RICH (ACIDIC).  
 CC FT DOMAIN 510 545 ASP/GLU-RICH (ACIDIC).  
 CC SQ SEQUENCE 606 AA; 66407 MW; 63D0EE551D48E CRC64;

Query Match 17.4%; Score 149.5; DB 1; Length 606;  
 Best Local Similarity 25.2%; Pred. No. 0.023;  
 Matches 41; Conservative 22; Mismatches 47; Indels 53; Gaps 4;  
 16 GGSLRGNPSS-----QVDEQMNVRVEEQQOQLRQEEHTARNGEVEEPGQND 71

DB 390 GGGANNTTTSFKSGEEREEEREEEREEEREEEREEEREEEREE 445  
 QY 72 SQOQLENNRFTSVDESSGNGOEDEDE----- 103  
 DB 446 GEEVEVEEBS-----DESEEREEEREEEREEEREEEREEEREE 439  
 QY 104 -----HAGEODEDEDEEMDQESDDPQDSDSSRED 135  
 DB 500 EEAQCEPTLHLEGGEDSDSDDEERDEDEERDEDEDDDDDD 542

RESULT 5  
 IF38\_SCHPO STANDARD; PRT; 918 AA.  
 ID IF38\_SCHPO  
 AC Q14164; Q13885; P78791; Q9P6P4;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Probable eukaryotic translation initiation factor 3 93 kDa subunit (eIF3 p93).  
 GN IF33 OR SPAC4A8.16C OR SPAC823.01C OR SPAC1E11.01C.  
 OS Schizosaccharomyces pombe (Fission Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 OC Schizosaccharomyces.  
 NC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Raylstrand M.A., Lyne M., Lyne R., Stewart A., Sgouras J., Peat N., Hayes L., Baker S., Baaham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Shelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Gymnopreux B., Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabell C., Fuchs M., Fritsch C., Holzer E., Moesli D., Hilbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wandut R., Fumelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Kottler S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revelante J.L., Moreno S., Armstrong J., Forburg S.L., Cerritelli L., Lowe T., McCombie M.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880 (2002).  
 RN [2]  
 RP SEQUENCE OF 571-918 FROM N.A.  
 RC STRAIN=PR745;  
 RX MEDLINE=98162722; PubMed=9501991;  
 RA Yoshioke S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in Schizosaccharomyces pombe cDNAs";  
 RL DNA Res. 4:363-368 (1997).  
 CC -1- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF METHIONYL-TRNAI AND MRNA (BY SIMILARITY).  
 CC -1- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY SIMILARITY).  
 CC -1- DOMAIN: CONTAINS 1 PCI DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE EIF38 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -



NAB3 YEAST  
ID NAB3 YEAST STANDARD; PRT; 802 AA.  
AC P38956;  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Nucleic acid polyadenylated RNA-binding protein NAB3.  
GN NAB3 OR YPL190C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Saccharomycetes; Saccharomycetaceae; Saccharomycetes;  
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_Taxid=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Wilson S.M., Oberdorf A.M., Datar K.V., Swedlow J.R., Paddy M.R.,  
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA STRAIN=S288c / AB972;  
RX MEDLINE=97313271; PubMed=9169875;  
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,  
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,  
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,  
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,  
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,  
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,  
RA Hall J., Hebl U., Heumann K., Hilbert H., Hillier L.,  
RA Hunkeler-Smith S., Hyman R., Johnston M., Kaiman S., Kleene K.,  
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,  
RA Marathe R., Messenguy F., Mewes H.-W., Mittlepat S., Moestl D.,  
RA Mueller-Auer S., Namath A., Neutwich U., Oefner P., Pearson D.,  
RA Petel F.X., Pohl T.M., Purrelle D., Schafer M., Scharfe M.,  
RA Scherens B., Schramm S., Schroeder M., Sidic A.M., Tettelin H.,  
RA Urtreparatu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,  
RA Walsh S.V., Wambolt R., Wang Y., Wedler B., Wedler H., Winnett E.,  
RA Zhong W.W., Zollner A., Vo D.H., Han J.,  
RL "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."  
RT Nature 387:103-105(1997).  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: U05314; AAA81910.1; -  
CC EMBL: 273546; CA97903.1; -  
CC PIR: S48529; S48529.  
CC  
CC SGD: S0006111; NAB3.  
CC  
CC GO: GO:0005654; C:nucleoplasm; IDA.  
CC GO: GO:0008143; F:poly(A) binding activity; IDA.  
CC GO: GO:0006357; P:regulation of transcription from Pol II pro. .; IGI.  
CC InterPro: IPR000504; RNA\_rec\_mot.  
CC  
CC Pfam: PF00076; rrm\_1.  
CC SMART: SM00360; RRM, 1.  
CC PROSITE: PS00102; RRM, 1.  
CC DR RNA-binding; Nuclear protein.  
CC  
CC DOMAIN 330 401 RNA-BINDING (RRM).  
FT DOMAIN 68 71  
FT DOMAIN 87 93 POLY-GLU.  
FT DOMAIN 101 106 POLY-GLU.  
FT DOMAIN 108 115 POLY-ASP.  
FT DOMAIN 116 127 POLY-GLU.  
FT DOMAIN 128 137 POLY-ASP.  
FT DOMAIN 137 143 POLY-GLU.  
FT DOMAIN 603 608 POLY-GLN.  
FT DOMAIN 644 648 POLY-PRO.  
FT DOMAIN 698 703 POLY-GLN.

FT DOMAIN 723 728 POLY-GLN.  
FT DOMAIN 765 768 POLY-PRO.  
FT DOMAIN 769 784 POLY-GLN.  
SQ SEQUENCE 802 AA; 90438 MW; CD6D2C7F24A44993 CRC64;  
Query-Match 17.7%; Score 152.5; DB 1; Length 802;  
Best Local Similarity 28.8%; Pred. No. 0.02;  
Matches 42; Conservative 29; Mismatches 48; Indels 27; Gaps 6;  
QY 6 LSVGSKRRRTGSLRGNPSS-----SQVDEQRNRYVVEEQOOL-----RQEEF 51  
DB 18 LSVDSNSNE--NELMNNSSADDGIEFDAPEEERAREERENRQHEIDVNDDEEDKER 75  
QY 52 HTKRNQGVGVVERPGQNDSSQCGOLENNNRFTSVDEDSGNGRQEDDEHAGEDDEE 111  
DB 76 KGEENGVEINTE-----EEEEHQQKGGADD---DDDNNEEEEDDDDDDDDDDD 126  
QY 112 DEEEEMDQE--SDPDFQSDSSRED 135  
DB 127 DEEEEEEEREGDNGSVGSDSAED 152

## RESULT 3

IE68 HSVSA STANDARD; PRT; 407 AA.

AC Q01042;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-APR-1993 (Rel. 25, Last annotation update)  
DE Immediate-early protein.  
GN 73 OR ECUF1.  
OS Herpesvirus saimiri (strain 11).  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Gammaherpesvirinae; Rhadinovirus.  
OX NCBI\_Taxid=10383;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92333688; PubMed=1321287;  
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,  
RA Newman C., Wiltman S., Craxton M.A., Coleman H., Fleckenstein B.,  
RA Honess R.W.;  
RT "Primary structure of the herpesvirus saimiri genome."  
RL J. Virol. 66:5047-5058(1992).

RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92230228; PubMed=1314457;  
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;  
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of  
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic  
RT organization between HVS and Epstein-Barr virus."  
RL Virology 188:296-310(1992).  
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUP TOGETHER HSV-1 AND  
CC HSV-2 IE-68 (US1), EBV-1 65, EBV-4 (ORF4), PRV RBP40, AND VZV 63.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X64346; CAA45696.1; -  
CC EMBL: M66409; AAA46149.1; -  
CC EMBL: S76368; AAB21116.1; -  
CC Early protein.  
CC  
CC DOMAIN 60 241 GLU-RICH (ACIDIC).  
SQ SEQUENCE 407 AA; 46617 MW; FPD399CAB82CE136C CRC64;

Query Match 17.4%; Score 150; DB 1; Length 407;  
Best Local Similarity 24.8%; Pred. No. 0.015;  
Matches 35; Conservative 32; Mismatches 64; Indels 10; Gaps 2;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 19:54:00 ; Search time 18 Seconds

(without alignments)  
436.303 Million cell updates/sec

Title: US-10-245-618-4

Sequence: 1 MNGELSLVSGKRRRTGGSLR.....STVDLPVHQLSPPTTKTK 167

Scoring table: BIOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Match Length	DB ID	Description
1	160.5	18.7	678 1	GARP_PLAUF
2	152.5	17.7	802 1	NAB3_YEAST
3	150	17.4	407 1	IE68_HYSA
4	149.5	17.4	606 1	CENB-CRIGR
5	149.5	17.4	918 1	IF38_SCHPO
6	147	17.1	764 1	UBF1_HUMAN
7	146.5	17.0	1121 1	MYT1_HUMAN
8	145.5	16.9	1489 1	YGPO_YEAST
9	145	16.9	738 1	SS21_RAT
10	144.5	16.8	411 1	MP62_LYPI
11	141.5	16.5	764 1	UBF1_RAT
12	141.5	16.5	2492 1	ATRX_HUMAN
13	139.5	16.2	765 1	UBP1_MOUSE
14	138.5	16.1	1394 1	CNG4_BOVIN
15	138	16.0	599 1	CENB_HUMAN
16	133.5	15.5	1719 1	PRD2_HUMAN
17	132.5	15.4	752 1	DRS1_YEAST
18	132.5	15.4	1972 1	BAB2_HUMAN
19	132	15.3	396 1	TRT_DROME
20	132	15.3	599 1	CENB_MOUSE
21	132	15.3	602 1	SPT8_YEAST
22	132	15.3	1099 1	NGX1_HUMAN
23	131.5	15.3	797 1	VG48_HYSA
24	131	15.2	676 1	SS21_HUMAN
25	131	15.2	713 1	NUCL_MESAU
26	130.5	15.2	110 1	THYA_MOUSE
27	130.5	15.2	701 1	UBP2_XENLA
28	130.5	15.2	2476 1	ATRX_MOUSE
29	130	15.1	482 1	T2EA_YEAST
30	130	15.1	694 1	NUCL_CHICK
31	129.5	15.1	739 1	DAXX_MOUSE
32	128.5	14.9	110 1	THYA_HUMAN
33	128.5	14.9	427 1	EBP2_YEAST

34	128.5	14.9	852 1	SRCH_RABIT	P16230	oryctolagus
35	128	14.9	712 1	NUCL_RAT	P13383	rattus norv
36	127.5	14.8	675 1	SG1_RAT	O35134	rattus norv
37	127	14.8	109 1	THYA_BOVIN	P01252	bos taurus
38	126.5	14.7	239 1	CENB_SHEEP	P49451	ovis aries
39	126	14.7	600 1	ABRA_PLAUF	P22620	plasmodium
40	125	14.7	743 1	ABRA_PLAUF	P22620	plasmodium
41	125.5	14.6	590 1	YM72_YEAST	O05021	saccharomyc
42	125.5	14.6	639 1	GLCX_SOYBN	P11827	glycine max
43	125	14.5	300 1	NGX1_BISBI	O46383	bison bison
44	125	14.5	358 1	CBF1_KLUYA	P49379	kluyveromyc
45	125	14.5	544 1	NFL_XENLA	P35616	xenopus lae

## ALIGNMENTS

RESULT 1	GARP_PLAUF	STANDARD;	PRT;	678 AA.
AC	P13816;			
DT	01-JAN-1990 (Rel. 13, Created)			
DT	01-JAN-1990 (Rel. 13, Last sequence update)			
DT	15-JUN-1999 (Rel. 38, Last annotation update)			
DE	Glutamic acid-rich protein precursor.			
GN	GARP.			
OS	Plasmodium falciparum (isolate FC27 / Papua New Guinea).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5837;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89040048; PubMed=2903445;			
RA	Trigita T., Stahl H.-D., Creweher P.E., Silva A., Anders R.F.,			
RA	Kemp D.U.;			
RT	"Structure of a Plasmodium falciparum gene that encodes a glutamic			
RT	acid-rich protein (GARP)."			
RL	Mol. Biochem. Parasitol. 31:199-202(1988).			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL, J03998; AAA39605.1; -			
DR	PIR; A54514; A54514.			
KW	Repeat; Malaria; Antigen; Signal.			
FT	SIGNAL	1	25	
FT	CHAIN	26	678	
FT	DOMAIN	120	164	
FT	DOMAIN	372	416	
FT	DOMAIN	417	441	
FT	DOMAIN	576	604	
FT	DOMAIN	605	653	
FT	DOMAIN	654	663	
SO	SEQUENCE	678 AA;	80551 MW;	2A8F85606496E8A9E CRC64;
Query Match		18.7%;	Score 160.5;	DB 1; Length 678;
Best Local Similarity		27.3%;	Pred. No. 0.0058;	
Matches	30;	Conservative	31;	Mismatches 34; Indels 15; Gaps 1;
QY	27 QVDEDEMRVREEDQQLRQDEEHRTANGEVGVPRPGQNDSSQQLLENNRFTS 86			
DB	566 QDEEVEEVEDEEEEEEDEEE-----EEEEEEEEEEDEDEDE 610			
QY	87 VVEDSSGNOEQEEDENHAGDEDEDEEENDQSDPDSRDE 136			
DB	611 DEDDAEDDDAEEDDDAEEDDDDEDEDEDEDEDEDEDEDEDEDE 660			

RESULT 2

**THIS PAGE BLANK (uspto)**

Nature 344, 830-836, 1990

A/Title: Nucleolar transcription factor hNBF contains a DNA-binding motif with homology

A/Reference number: S09318; MUID:90231434; PMID:2330041

A/Accession: S09318

A/Molecule type: mRNA

A/Residues: 1-764 <JAN>

A/Cross-references: EMBL:X53461; NID:g37573; PIDN:CA37548.1; PID:g37574

A/Note: part of this sequence was confirmed by amino acid sequencing

C/Superfamily: Unassigned HMG box proteins; HMG box homology

C/Keywords: DNA binding; transcription regulation

F/109-184/Domain: HMG box homology <HMG1>

F/193-268/Domain: HMG box homology <HMG2>

F/294-366/Domain: HMG box homology <HMG3>

F/404-479/Domain: HMG box homology <HMG4>

F/482-553/Domain: HMG box homology <HMG5>

F/564-638/Domain: HMG box homology <HMG6>

Query Match

Best Local Similarity 17.1%; Score 147; DB 2; Length 764;

Matches 41; Conservative 35; Mismatches 57; Indels 44; Gaps 5;

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

JC5113

ribosomal transcription factor UBF2 - Chinese hamster

C/Species: Crictetus griseus (Chinese hamster)

C/Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 12-Feb-1999

C/Accession: JC5113

R/Bolivar, J.; Goenechea, L.G.; Grenett, H.; Pendon, C.; Valdivia, M.M.

Gene 176, 257-258, 1996

A/Title: Cloning and sequencing of the genes encoding the hamster ribosomal transcription

A/Reference number: JC5112; MUID:97075939; PMID:8918262

A/Accession: JC5113

A/Molecule type: mRNA

A/Residues: 1-727 <BOU>

A/Cross-references: GB:L42571

C/Comment: This factor binds to the ribosomal RNA gene's promoter and forms a stable pre

C/Superfamily: Unassigned HMG box proteins; HMG box homology

F/109-184/Domain: HMG box homology <HMG1>

F/257-329/Domain: HMG box homology <HMG3>

F/445-516/Domain: HMG box homology <HMG5>

Query Match

Best Local Similarity 16.9%; Score 145.5; DB 2; Length 727;

Matches 40; Conservative 37; Mismatches 56; Indels 45; Gaps 5;

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Search completed: December 10, 2003, 19:59:50

Job time : 22 secs

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

Qy

Db

RESULT 15

A: Molecule type: DNA  
A: Residues: 1-1094 <WF2>

C/Accession: S09318





A;Residues: 1-409 &lt;STO&gt;



**THIS PAGE BLANK (USPTO)**

	Query Match Similarity	19.0%	Pred. 163.5,	DB 21,	Length 1162;
	Best Local Similarity	29.1%	Pred. No. 7,6e-06;		
	Matches	32;	Conservative	26;	Mismatches 35; Indels 17; Gaps 14.
Qy	27 QVDEEQMNRVVVEEQQOQLROCEEHTARNGEVGVPRPGQNSOGGLENNRRPTLS	86			
Dd	636 QODEQQQDEQQQDDEQQDDEQQDD-----QDDEQQQDEQQDDEQQD	678			
Qy	87 VDEDSGNQEQAEDDEHAGQDDEBDEEEEMWDCSDDPGDSDSRDEE	136			
Dd	679 EQQDDEQQDDEQQDDEQQDDEQQDDEQQDDEQQDDEQQDDEQQDDEQQDDE	728			

RESULT 15	
AAVS8500	
ID	AAVS8500 standard; Protein; 1162 AA
XX	
AC	AAVS8500;

PT New methods and compositions for the detection of human herpesvirus -  
XX  
PS Claim 2; Page 59-62; 68pp; English.  
XX

CC Sequences AAY58480-Y58532 represent immunogenic polypeptides derived  
CC from human herpes virus type 8 (HHV8 a gammaherpesvirus). HHV8 plays an  
CC important role in the pathogenesis of AIDS-related Kaposi's sarcoma. The  
CC invention relates to a novel method of detecting the presence of human  
CC herpesvirus 8 in a biological sample using peptides representative of  
CC dominant antigenic regions of HHV8. The method comprises contacting one  
CC or more isolated, immunogenic HHV8 peptides with an antibody-containing  
CC biological sample, and detecting the formation of a complex between the  
CC peptide and the antibody. The presence of a peptide-antibody complex  
CC indicates the presence of human herpesvirus 8. The detection of HHV8  
CC infection can be used to diagnose AIDS-associated Kaposi's sarcoma. The  
CC HHV8-specific antibodies are useful therapeutically when for the passive  
CC immunisation of a human against HHV8 infection, thereby reducing HHV8  
CC related disease. The detection assays are highly specific, sensitive and  
CC accurate. Early detection and treatment of Kaposi's sarcoma could  
CC diminish the severity of symptoms related to AIDS and the sensitive  
CC techniques could reduce erroneous characterisations of skin disorders.  
CC Previous assays for HHV8 antibodies such as immunofluorescence assays,  
CC immunoblots and enzyme immunoassays lack the sensitivity and accuracy  
CC needed for reliable diagnosis of Kaposi's sarcoma. Further advantages  
CC of the assays are that reproducible results are obtained and the method  
CC is suitable for rapid throughput and screening of samples economically.

[illegible]

Search completed: December 10, 2003, 19:57:24  
Job time : 41 secs

[illegible]

AC	AAY96255;
XX	
DT	11-SEP-2000 (first entry)
DE	Kaposi's sarcoma-associated herpesvirus LANA.
XX	
KW	Kaposi's sarcoma-associated herpesvirus; KSHV; rhadno virus;
KW	latency-associated nuclear antigen; LANA; gamma-2 herpes virus;
KM	Human herpes virus 8; HHV8; rhadno virus cis-acting element; RVCAE;
KW	Kaposi's sarcoma; primary effusion lymphoma; PEL;
KW	human immunodeficiency virus; HIV; multicentric Castleman's disease.
OS	Kaposi's sarcoma-associated herpesvirus.
XX	
FH	Key Location/Qualifiers
FT	Domain 14..17
FT	/note= "nuclear localisation signal, NLS"
FT	Domain 64..70
FT	/note= "nuclear localisation signal, NLS"
FT	Region 320..429
FT	/note= "acidic repeat region"
FT	Region 430..549
FT	/note= "Gln, Glu, Pro-rich region"
FT	Region 550..589
FT	/note= "Gln, Glu, Pro, Arg-rich region"
FT	Region 590..759
FT	/note= "Gln, Glu, Asp-rich region"
FT	Region 760..840
FT	/note= "Gln, Glu-rich region"
PN	MO200029626-A1.
XX	
PD	25-MAY-2000.
XX	
PZ	19-NOV-1999; 99WO-US27508.
XX	
PR	19-NOV-1998; 98US-0109422.
PR	21-APR-1999; 99US-0298568.
PA	(KIEF/) KIEFF E D.
PA	(BALL/) BALLESTAS M E.
PA	(KAYE/) KAYE K M.
PT	Kieff ED, Ballestas ME, Kaye KM;
XX	
DR	WPI; 2000-387829/33.
DR	N-PSDB; AAA30290.
XX	
PT	Treating or preventing a disease associated with rhodno virus
PT	Infection in a mammal which includes Kaposi's Sarcoma and Primary
PT	Effusion Lymphoma -
XX	
PS	Disclosure; Fig 7; 70pp; English.
XX	
CC	The present sequence is the Kaposi's sarcoma-associated herpesvirus,
CC	(KSHV) latency-associated nuclear antigen (LANA). KSHV is also known
CC	as Human Herpes Virus 8 (HHV8) and belongs to the rhadno virus, or
CC	gamma-2 herpes virus class. The LANA protein is necessary for the
CC	efficient persistence of rhadno virus DNA in mammalian cells. Persistent
CC	rhadno virus infection is implicated in a variety of diseases e.g.
CC	Kaposi's Sarcoma (KS), Primary Effusion Lymphoma (PEL), and multicentric
CC	Castleman's disease. In addition, KS is a common malignancy in HIV
CC	patients. KSHV persists in host cells in a latent form. One of the few
CC	genes expressed from the latent viral DNA is LANA. LANA associates with
CC	both human chromosomes and with the rhadno virus cis-acting element
CC	(RVCAE), thereby providing a tethering function: the KSHV DNA episome is
CC	"tied" to the host chromosomes. This allows the viral DNA to persist in
CC	the host cell. The present sequence may be used to screen and identify
CC	molecules that inhibit LANA interaction with RVCAE, thereby interfering
CC	with the latency cycle of this virus. Potential antiviral treatments for
CC	the above mentioned diseases may therefore be based on LANA deregulation
XX	
SQ	Sequence 1162 AA;

AC ABG11270;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #11261.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 XX  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSEQ-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI: 2001-639362/73.  
 XX  
 DR N-PSDB; AAS75457.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID No 41629; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 167 AA;

Query Match 19.2%; Score 165.5; DB 22; Length 167;  
 Best Local Similarity 26.0%; Pred. No. 5.3e-07;  
 Matches 34; Conservative 32; Mismatches 44; Indels 21; Gaps 2;  
 QY 10 SKRRRTGSLRGNPSSQVDEQMNRVVEEQOQLRQOEHTANGEVGVPRPGQ 69  
 DB 57 SLRR-----RKEGEEEEEEEEEEEEEEEEEEEEEEEE-----E 95  
 QY 70 NDSGQGLLENNNRFTSVDESSGQOEQDEBEHAGQODEDEDEEENDQSDPQSD 129  
 DB 96 EEEEEEIEEGEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 155  
 QY 130 DSSREDEHTT 140  
 DB 156 EEEEEEEAKT 166

RESULT 12  
 AAM03627  
 ID AAM03627 standard; peptide; 360 AA.  
 XX  
 AC AAM03627;  
 XX  
 DT 25-MAR-2003 (updated)  
 DT 17-DEC-1996 (first entry)  
 XX  
 DE Human follicle stimulating hormone GPR N-terminal sequence.  
 XX  
 KM G-protein coupled receptor; ligand binding assay; transmembrane domain;  
 KM schizophrenia; dopamine; cAMP; adenosine; thrombin; adrenergic; opsin;  
 KM muscarinic acetylcholine; endothelin; bombesin; endocrine; rhodopsin;  
 KM odorant; cytomagalovirus; serotonegic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US5508384-A.  
 XX  
 PD 16-APR-1996.  
 XX  
 PF 09-SEP-1993; 93US-0118270.  
 XX  
 PR 09-SEP-1993; 93US-0118270.  
 PR 10-SEP-1992; 92US-0943236.  
 XX  
 PA (UYNV) UNIV NEW YORK STATE.  
 XX  
 PI Murphy RB, Schuster DI;  
 XX  
 DR WPI: 1996-208785/21.  
 XX  
 PT New dopamine receptor peptide - useful as antipsychotic agent, e.g.  
 PT for treating schizophrenia  
 XX  
 PS Disclosure; Fig 8B(2); 184pp; English.  
 XX  
 CC Proteins AAM02657-W02720 represent a range of G-protein coupled receptor  
 CC (GPR) proteins selected from cAMP, adenosine, muscarinic acetylcholine,  
 CC adrenergic, thrombin, endothelin, bombesin, endocrine, rhodopsin, opsin,  
 CC odorant, cytomagalovirus and other GPR proteins. The peptides  
 CC AAM03578-W03651 represent the N-terminal fragments of the above  
 CC proteins. The receptor proteins were used to design polypeptides, pref.  
 CC based on the transmembrane domains, for use in G-protein coupled receptor  
 CC ligand binding assays. The polypeptide fragments retain biological  
 CC activity such as binding a GPR ligand or modulating GPR ligand binding to  
 CC a GPR (see AAM02747-W02899 for examples of polypeptide fragments).  
 CC The polypeptide fragments can be used in compositions for treating  
 CC subjects suffering from a pathology related to a GPR abnormality e.g. a  
 CC psychotic disorder such as schizophrenia.  
 CC (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 360 AA;

Query Match 19.2%; Score 165; DB 17; Length 360;  
 Best Local Similarity 26.3%; Pred. No. 1.4e-06;  
 Matches 31; Conservative 33; Mismatches 54; Indels 0; Gaps 0;  
 QY 29 DEQMNRVVEEQOQLRQOEHTANGEVGVPRPGQNDGQGLLENNNRFTSV 88  
 DB 169 EEE 248  
 QY 89 EDSGQGLLENNNRFTSVDESSGQOEQDEBEHAGQODEDEDEEENDQSDPQSD 146  
 DB 249 EEE 306  
 RESULT 13  
 AAM03626  
 ID AAM03626 standard; peptide; 412 AA.

PN WO200277036-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US08614.  
 XX 21-MAR-2001; 2001US-277705P.  
 XX (LEUNG/) LEUNG D W.  
 XX Leung DW, Bergman PA, Lofquist A, Pietz GE, Tompkins CK;  
 XX Waggoner DW;  
 XX WPI; 2003-058367/05.  
 XX  
 XX Producing monodispersed preparation of polyanionic polymer for therapy,  
 XX by expressing vector comprising ligation product of oligonucleotides  
 XX encoding glutamate/aspartate residues in host cell and isolating the  
 XX product -  
 XX  
 XX Example 6; Page 41; 74pp; English.  
 XX  
 XX The present invention describes a method (M) for producing a  
 XX monodispersed preparation of a polyanionic polymer (PP) larger than  
 XX 10 kD. (M) involves inserting into an expression vector (EV) a ligation  
 XX product formed by ligating together oligonucleotides that encode  
 XX glutamate/aspartate residues, expressing EV in a host cell, and  
 XX isolating the protein product. (P) of EV, where (P) is PP and at least  
 XX 80% of PP is approximately of the same molecular weight. Also described:  
 XX (1) a recombinant fusion protein (I) comprising a polyanionic polypeptide  
 XX and another polypeptide at either one end or at both ends of it; (2) a  
 XX polyanionic polymer (II) conjugate comprising a polyanionic polymer and  
 XX leucine, where the polyanionic polymer is polyglutamic acid or  
 XX polyaspartic acid; (3) a vector (III) comprising a cassette which  
 XX comprises a nucleotide sequence encoding a polyanionic polymer and at  
 XX least one other nucleotide sequence, where the polyanionic polymer is  
 XX polyglutamic acid or polyaspartic acid; (4) production of (I); (5) a cell  
 XX (IV) comprising (III) or a vector that comprises a nucleotide sequence  
 XX that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a  
 XX recombinantly-produced polyanionic polymer (V) that is of any molecular  
 XX weight or is larger than 10 kD, and is conjugated to another protein.  
 XX (I) is useful for treating a disease or ailment in an individual by  
 XX administering (I) to the individual. (I) is also useful for delivering an  
 XX effective amount of a pharmaceutically active agent, a therapeutic  
 XX protein or a drug to a patient in need of it, or for diagnostic and  
 XX testing or research purposes. AB222045 to AB222131 and ABP56374 to  
 XX ABP6400 represent sequences used in the exemplification of the present  
 XX invention.  
 XX  
 XX Sequence 187 AA;  
 XX  
 XX Query Match 19.4%; Score 167; DB 24; Length 187;  
 XX Best Local Similarity 27.2%; Pred. No. 4.5e-07;  
 XX Matches 31; Conservative 31; Mismatches 52; Indels 0; Gaps 0;  
 XX  
 XX 23 PSSQVDEQNMNRVVEEQOOLRQOEHTNARNGVGVPRPGQNDGQGLLENNN 82  
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 XX 4 PEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 63  
 XX  
 XX 83 RFIIVDEDSGNQOEDEEHAGQDEDEDEEEMDQESDDPQSDSSRDE 136  
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 XX 64 EEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 117  
 XX  
 XX RESULT 10  
 XX ABP56380  
 XX ID ABP56380 standard; Peptide; 198 AA.  
 XX  
 XX ABP56380;  
 XX  
 XX 11-MAR-2003 (first entry)  
 XX  
 XX Polyanionic fusion protein product #1.  
 XX

XX  
 XX Polyanionic polymer; bioactivity; water solubility.  
 XX Synthetic.  
 XX WO200277036-A2.  
 XX 03-OCT-2002.  
 XX 21-MAR-2002; 2002WO-US08614.  
 XX 21-MAR-2001; 2001US-277705P.  
 XX (LEUNG/) LEUNG D W.  
 XX Leung DW, Bergman PA, Lofquist A, Pietz GE, Tompkins CK;  
 XX Waggoner DW;  
 XX WPI; 2003-058367/05.  
 XX  
 XX Producing monodispersed preparation of polyanionic polymer for therapy,  
 XX by expressing vector comprising ligation product of oligonucleotides  
 XX encoding glutamate/aspartate residues in host cell and isolating the  
 XX product -  
 XX  
 XX Disclosure; Page 28; 74pp; English.  
 XX  
 XX The present invention describes a method (M) for producing a  
 XX monodispersed preparation of a polyanionic polymer (PP) larger than  
 XX 10 kD. (M) involves inserting into an expression vector (EV) a ligation  
 XX product formed by ligating together oligonucleotides that encode  
 XX glutamate/aspartate residues, expressing EV in a host cell, and  
 XX isolating the protein product (P) of EV, where (P) is PP and at least  
 XX 80% of PP is approximately of the same molecular weight. Also described:  
 XX (1) a recombinant fusion protein (I) comprising a polyanionic polypeptide  
 XX and another polypeptide at either one end or at both ends of it; (2) a  
 XX polyanionic polymer (II) conjugate comprising a polyanionic polymer and  
 XX leucine, where the polyanionic polymer is polyglutamic acid or  
 XX polyaspartic acid; (3) a vector (III) comprising a cassette which  
 XX comprises a nucleotide sequence encoding a polyanionic polymer and at  
 XX least one other nucleotide sequence, where the polyanionic polymer is  
 XX polyglutamic acid or polyaspartic acid; (4) production of (I); (5) a cell  
 XX (IV) comprising (III) or a vector that comprises a nucleotide sequence  
 XX that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a  
 XX recombinantly-produced polyanionic polymer (V) that is of any molecular  
 XX weight or is larger than 10 kD, and is conjugated to another protein.  
 XX (I) is useful for treating a disease or ailment in an individual by  
 XX administering (I) to the individual. (I) is also useful for delivering an  
 XX effective amount of a pharmaceutically active agent, a therapeutic  
 XX protein or a drug to a patient in need of it, or for diagnostic and  
 XX testing or research purposes. AB222045 to AB222131 and ABP56374 to  
 XX ABP6400 represent sequences used in the exemplification of the present  
 XX invention.  
 XX  
 XX Sequence 198 AA;  
 XX  
 XX Query Match 19.4%; Score 167; DB 24; Length 198;  
 XX Best Local Similarity 27.2%; Pred. No. 4.8e-07;  
 XX Matches 31; Conservative 31; Mismatches 52; Indels 0; Gaps 0;  
 XX  
 XX 23 PSSQVDEQNMNRVVEEQOOLRQOEHTNARNGVGVPRPGQNDGQGLLENNN 82  
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 XX 11 PEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 70  
 XX  
 XX 83 RFIIVDEDSGNQOEDEEHAGQDEDEDEEEMDQESDDPQSDSSRDE 136  
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |  
 XX 71 EEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 124  
 XX  
 XX RESULT 11  
 XX ABG11270  
 XX ID ABG11270 standard; Protein; 167 AA.  
 XX

[illegible]

XX	FR	21-MAR-2001; 2001US-277705P.
XX	PA	(LEUNG/) LEUNG D W.
XX	PI	Leung DW, Bergman PA, Lofquist A, Pietz GE, Tompkins CK,
XX	PI	Magnoner DW;
XX	DR	WP1; 2003-058367/05.
XX	PT	Producing monodispersed preparation of polyanionic polymer for therapy,
XX	PT	by expressing vector comprising ligation product of oligonucleotides
XX	PT	encoding glutamate/aspartate residues in host cell and isolating the
XX	PT	product -
XX	PS	Example 4; Page 35; 74pp; English.
XX	CC	The present invention describes a method (M) for producing a
XX	CC	monodispersed preparation of a polyanionic polymer (PP) larger than
XX	CC	10 kD. (M) involves inserting into an expression vector (EV) a ligation
XX	CC	product formed by ligating together oligonucleotides that encode
XX	CC	glutamate/aspartate residues, expressing EV in a host cell, and
XX	CC	isolating the protein product (P) of EV, where (P) is PP and at least
XX	CC	80% of PP is approximately of the same molecular weight. Also described:
XX	CC	(1) a recombinant fusion protein (I) comprising a polyanionic polypeptide
XX	CC	and another polypeptide at either one end or at both ends of it; (2) a
XX	CC	polyanionic polymer (II) conjugate comprising a polyanionic polymer and
XX	CC	leukine, where the polyanionic polymer is polylutamic acid or
XX	CC	polyaspartic acid; (3) a vector (III) comprising a cassette which
XX	CC	comprises a nucleotide sequence encoding a polyanionic polymer and at
XX	CC	least one other nucleotide sequence, where the polyanionic polymer is
XX	CC	polylutamic acid or polyaspartic acid; (4) production of (I); (5) a cell
XX	CC	(IV) comprising (III) or a vector that comprises a nucleotide sequence
XX	CC	that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a
XX	CC	recombinantly-produced polyanionic polymer (V) that is of any molecular
XX	CC	weight or is larger than 10 kD, and is conjugated to another protein.
XX	CC	(I) is useful for treating a disease or ailment in an individual by
XX	CC	administering (I) to the individual. (I) is also useful for delivering an
XX	CC	effective amount of a pharmaceutically active agent, a therapeutic
XX	CC	protein or a drug to a patient in need of it, or for diagnostic and
XX	CC	testing or research purposes. AB222045 to AB222131 and ABP6374 to
XX	CC	ABP6400 represent sequences used in the exemplification of the present
XX	CC	invention.
XX	SQ	Sequence 186 AA;
XX		
XX	Query Match	19.4%; Score 167; DB 24; Length 186;
XX	Best Local Similarity	27.2%; Prid. No. 4,4e-07;
XX	Matches 31; Conservative 31; Mismatches 52; Indels 0; Gaps 0;	
XX	QY	23 PSSQGVVEOMNRVVEEEOQOLROOEHEHTANGVEVGVPRPGQNDQOQLSENNN 82
XX		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :
XX	DB	11 PEEEEEDEEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 70
XX		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :
XX	QY	83 RFIIVDESSGNQOEDEDEHAGQOEDEDEEEEMDQESDDPDQSDSSREDE 136
XX		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :
XX	DB	71 EEEEEEDEEEEEEDEEEEEEDEEEEEEDEEEEEEDEEEEEEDEEEEEEDEEEEEE 124
XX		: : : : :   : : : : :   : : : : :   : : : : :   : : : : :
XX	RESULT 9	
XX	ABP6384	
XX	ID	ABP6384 standard; Peptide; 187 AA.
XX	AC	ABP6384;
XX	DT	11-MAR-2003 (first entry)
XX	DE	Polyanionic fusion protein product #4.
XX	KW	Polyanionic polymer; bioactivity; water solubility.
XX	OS	Synthetic.
XX	XX	



PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS75467.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 XX  
 PS Claim 20; SEQ ID NO 41639; 103bp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 \*SQ Sequence 254 AA;  
 Query Match 20.3%; Score 174.5; DB 22; Length 254;  
 Best Local Similarity 25.0%; Pred. No. 1.4e-07;  
 Matches 39; Conservative 38; Mismatches 44; Indels 35; Gaps 4;  
 QY 9 GSKRRRTGSLRGNPSSQVDEQNMNVVEEQOQLRQOEHTANNGEVVGPVPPGG 68  
 DB 55 GEERRRK-EKKGEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 95  
 QY 69 QNDSQOQLLENNRFTISVDESSGNOEEDDEHAGOEDEDEEEMQESDDPDQS 128  
 DB 96 --EEEEEEEEE-----E:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|: 146  
 QY 129 DSGREDEHTNTSVNSSIVD-----LPVQL 157  
 DB 147 EEEEEEGLSIKMTSTKTSLDINGFKLLPVPI 182  
 RESULT 4  
 AAR70491 standard; Protein; 562 AA.  
 AAR70491:  
 19-DEC-1995 (first entry)  
 Leucocytozoan protozoa structural protein epitope.  
 Leucocytozoan protozoa structural protein; epitope; vaccine; fowl;  
 leucocytozoanosis; treatment.  
 Leucocytozoan protozoa sp.  
 JPO7089995-A.  
 04-APR-1995.

XX 10-SEP-1993; 93JP-0226078.  
 XX  
 XX 10-SEP-1993; 93JP-0226078.  
 PR  
 XX  
 PA (DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.  
 PA (NIS-) NISSEIKEN KK.  
 XX  
 DR WPI; 1995-167252/22.  
 DR N-PSDB; AAQ87587.  
 XX  
 PT Immune inducing polypeptide against Leucocytozoan protozoa - useful  
 PT in production of vaccines for treatment of leucocytozoanosis in  
 PT fowl.  
 XX  
 PS Claim 1; Page 12-14; 20pp; Japanese.  
 XX  
 CC AAR70491-93 are polypeptides having a whole or partial epitope of a  
 CC structural protein of Leucocytozoan protozoa encoded by AAQ87587-89.  
 CC The polypeptides and DNA encoding them are useful in the production  
 CC of vaccines for the treatment of leucocytozoanosis of fowl.  
 XX  
 \*SQ Sequence 562 AA;  
 Query Match 20.1%; Score 172.5; DB 16; Length 562;  
 Best Local Similarity 27.0%; Pred. No. 5.3e-07;  
 Matches 40; Conservative 31; Mismatches 64; Indels 13; Gaps 2;  
 QY 2 NQELISGSKRRRTGSLRGNPSSQVDEQNMNVVEEQOQLRQOE-----EEN 52  
 DB 25 NNKIKVKEKEENBENEKEEEOEEOEIOEIOEIOEIOEIOEIOEIOEIOEIOEIOE 84  
 QY 53 TARNGEVVGVPVPPGGQNDSSQOQLLENNRFTISVDE-----EDSSGNOEEDDEHAGOE 108  
 DB 85 EEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 140  
 QY 109 DEEDDEEEMDQESDDPDQSDSSRDE 136  
 DB 145 NEEEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEEDEEE 172  
 RESULT 5  
 ABG05354 standard; Protein; 144 AA.  
 ABG05354:  
 13-FEB-2002 (first entry)  
 Novel human diagnostic protein #5345.  
 Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder.  
 Homo sapiens.  
 WO200175067-A2.  
 11-OCT-2001.  
 30-MAR-2001; 2001WO-US08631.  
 31-MAR-2000; 2000US-0540217.  
 23-AUG-2000; 2000US-0649167.  
 (HYSE-) HYSEQ INC.  
 Drmanac RT, Liu C, Tang YT;  
 WPI; 2001-639362/73.  
 N-PSDB; AAS69541.  
 New isolated polynucleotide and encoded polypeptides, useful in



PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 41624; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 194 AA;  
 Query Match 20.3%; Score 175; DB 22; Length 194;  
 Best Local Similarity 29.8%; Pred. No. 9.1e-08;  
 Matches 34; Conservative 30; Mismatches 42; Indels 8; Gaps 2;  
 Oy 23 PSSQVDEQNNRVVEEQOQLRQOEHTARNGEVGVPRPGGQNDQOGLLENNN 82  
 Db 28 PSEKEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEE 84  
 Oy 83 RFSIVDEDSGNOEEDSEHAGODEDEDEEMDQESDPQSDSDSREDE 136  
 Db 85 -----EEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEE 133  
 RESULT 2  
 ID ABG06966 standard; Protein; 254 AA.  
 AC ABG06966;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE  
 XX Novel human diagnostic protein #6957.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS71153.  
 RR

XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 20; SEQ ID No 37325; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 XX  
 SQ Sequence 254 AA;  
 Query Match 20.3%; Score 174.5; DB 22; Length 254;  
 Best Local Similarity 25.0%; Pred. No. 1.4e-07;  
 Matches 39; Conservative 38; Mismatches 44; Indels 35; Gaps 4;  
 Oy 9 GSKRRRTGSLRGNPSSQVDEQNNRVVEEQOQLRQOEHTARNGEVGVPRPG 68  
 Db 55 GEERRRR--EKGBSEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEE 95  
 Oy 69 QNDSQOQLLENNNRFSIVDEDSGNOEEDSEHAGODEDEDEEMDQESDPQ 128  
 Db 96 --EEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEE 146  
 Oy 129 DSSGREDEHTHNSVTNSSIVD-----LPHQL 157  
 Db 147 EEEEEEEEGSLIKYMTSTKTLSDINGFEKLLPVHPI 182  
 RESULT 3  
 ID ABG11280 standard; Protein; 254 AA.  
 AC ABG11280;  
 XX  
 DT 18-FEB-2002 (first entry)  
 DE  
 XX Novel human diagnostic protein #11271.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA  
 XX  
 PI  
 XX  
 DR  
 XX

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 19:52:45 ; Search time 40 Seconds

(without alignments)  
662.683 Million cell updates/sec

Title: US-10-245-618-4

Sequence: 1 MNGELSVGSKRRRTGSSLR.....SIYDLPHVQLSSPPYTKTK 167

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2001.DAT:\*  
23: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2002.DAT:\*  
24: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	20.3	194	22	ABG11265
2	174.5	20.3	234	22	ABG06566
3	174.5	20.3	254	22	ABG11280
4	172.5	20.1	562	16	AA070491
5	171	19.9	144	22	ABG05354
6	168.5	19.6	172	22	ABG11263
7	168	19.5	197	24	ABP56385
8	167	19.4	186	24	ABP56381
9	167	19.4	187	24	ABP56384

10	167	19.4	198	24	ABP56380	Polyanionic fusion
11	165.5	19.2	167	22	ABG11270	Novel human diago
12	165	19.2	360	17	AAW03627	Human follicle sti
13	164	19.1	412	21	AAW03626	Human thyrotropin
14	163.5	19.0	1162	17	AAW96255	Kaposi's sarcoma-a
15	163.5	19.0	1162	21	AAW58500	HRV8 ORF 73 protei
16	163.5	19.0	1162	21	AAW58500	Amino acid sequenc
17	163.5	19.0	1162	23	ABG05621	Kaposi's sarcoma-a
18	161	18.7	141	22	ABG04739	Novel human diago
19	161	18.7	181	24	ABP56382	Polyanionic fusion
20	160.5	18.7	1132	17	AAW97866	Chicken leucocyte
21	160	18.6	508	22	ABG04359	Novel human diago
22	159	18.5	186	22	ABG06961	Novel human diago
23	158	18.4	169	22	ABG06954	Novel human diago
24	158	18.4	199	22	ABG05352	Novel human diago
25	157.5	18.3	199	22	ABG05362	Novel human diago
26	157	18.3	217	22	ABG05366	Novel human diago
27	154	17.9	147	22	ABG11269	Novel human diago
28	153.5	17.8	1174	22	AAW85039	Novel human diago
29	152.5	17.7	802	23	ABG93215	S. cerevisiae BAX-
30	152	17.7	207	22	ABG05360	Shrimp white spot
31	151	17.6	235	22	ABG11274	Novel human diago
32	150.5	17.5	205	22	ABG03977	Novel human diago
33	150	17.4	682	21	AAW46506	Novel human diago
34	150	17.4	724	21	AAW46505	Arabidopsis thalia
35	150	17.4	778	21	AAW46504	Arabidopsis thalia
36	149.5	17.4	425	22	ABG06903	Novel human diago
37	149	17.3	231	22	ABG26553	Novel human diago
38	149	17.3	237	22	ABG06965	Novel human diago
39	149	17.3	237	22	ABG11276	Novel human diago
40	148.5	17.3	108	22	ABG05368	Novel human diago
41	148.5	17.3	165	22	ABG05368	Novel human diago
42	147.5	17.2	207	21	AAW4672	Gene 33 human secr
43	147	17.1	727	23	ABW97597	Novel human protei
44	147	17.1	745	22	ABG22102	Novel human diago
45	146	17.0	191	22	ABG03929	Novel human diago

ALIGNMENTS

RESULT 1	ABG11265	ABG11265 standard; Protein; 194 AA.
ID	ABG11265;	
AC	ABG11265;	
XX		
DT	18-FEB-2002 (first entry)	
XX		
DE	Novel human diagnostic protein #11256.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
XX		
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Dmanac RT, Liu C, Tang YT;	
XX		
DR	WPI: 2001-639362/73.	
XX	N-PSDB; AAS75452.	
PT	New isolated polynucleotide and encoded polypeptides, useful in	

**THIS PAGE BLANK (uspto)**

Job time : 5028 secs

---

KKIKHWYRADNFENKVALLLKHLKODATRSIVFVKRERVHELAETLRLAGINNC  
YLSEMGQIKRNGEIKRLTDGRVAVLVATAARGPDIIPVSHVINDFMRSGDTYH  
LREMGQAGKRGTAISVEAHDLILKIGRYIEBPKAVIDELRPTTAPSEKLTG  
KPSKVIYAKRAEKKKEKPKYKRYRDTKNIGKRRKSGTKMQDSSE"

gene  
6365..7102  
/gene="yfiC"  
/locus\_tag="t0268"  
6365..7102  
/gene="yfiC"  
/locus\_tag="t0268"  
/note="Corresponds to STY2835 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA067993.1"  
/db\_xref="GI:29136427"  
/translation="MSQSGVLRNNGFTFKQFVANDRCAMKVGTDGILGAPVAD  
VKRLIDIGTSGGLALMLAOTDSDVVDVDELDAQAMQAEVNAHSPHRIYHT  
DDIORMAPROTFEDLIISNPYEPGEVSTPQREARVATLIDHOTLLAIADCT  
EDGFECVLPQOIGNAFTQALNMGWHLRLRTDVAERAEALPHRVLLAFSPQAGECS  
DRIVRSGSDNYSESTALTOAFYLFM"

gene  
complement (7087..8709)  
/gene="nadB"  
/locus\_tag="t0269"  
complement (7087..8709)  
/gene="nadB"  
/locus\_tag="t0269"  
/note="Quinolinate synthetase B; corresponds to STY2834  
from Accession AL513382: Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="L-aspartate oxidase"  
/protein\_id="AA067994.1"  
/db\_xref="GI:29136428"  
/translation="MMTTPELSCDVLIIISGAAGLSIALRLAKHKVIVLSKPVSE  
STFAOGGIAVETEDSIAVSHEDTIIAGGICDRVAFVSNATCVOMLIDGV  
LFTHVQNGKESYHLRREGSHSRITLHAADATGKEVETLVSRAQNHVYIDERS  
NAVDLISDKIGLPGRVYVGVAMNNKEMVEKCAKSVIATGCAKSYOYTTND  
ISSGDIAMAMRACRIVANLBNOPHTALYHPOARFLLTEALRGSGAYLKPDPSR  
FMDVDVERGLAPDIIVARAIIDEMKRLGADCMFLDISHKPDVFRQHPMIYAKLID  
LMDLTKERIPIVPAAHYTCGVVVDYGRDVGALYIAEVSVTGIGHANRNASNL  
LECIYVMSAAMDIDRMPSVSHVSDVLPANDESVEVADRVVIOHNMHELRLIMDY  
VGIVRTKRLERALRRTITMLQOIDEVYAFVRSNNLELRNLVQVALLVRCAMRK  
ESRGLHTLDYVQQLAESPSILSPITPHNR"

gene  
9134..9709  
/gene="rpoB"  
/locus\_tag="t0270"  
9134..9709  
/gene="rpoB"  
/locus\_tag="t0270"  
/note="sigma-24; corresponds to STY2833 from Accession  
AL513382: Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="RNA polymerase sigma-E factor"  
/protein\_id="AA067995.1"

CDS  
7.58e-130 Length: 300431  
Score: 1831.00 Matches: 365  
Percent Similarity: 99.19% Conservative: 4  
Best Local Similarity: 98.12% Mismatches: 3  
Query Match: 98.12% Indels: 0  
Gaps: 0

US-09-921-992-50 (1-372) x AE016835 (1-300431)

Qy 1 MethisanginalaProleGlnArgArgSerThrArgIleTyValGlyAsnVal 20  
Db 83116 ATGCATTAACGAGGCTCCGATTCACGTAGAAATCGACACTATTTCGTTGGGAAATGTG 83175

Qy 21 ProileGlyAspGlyAlaProileAlaValGlnSerMetThrAsnThrArgThrThrAsp 40

|||||  
Db 83176 CCATTGGCGATGGCCCGCCCTATCGCGCTGCATGATGATCAACAGCGTACCATGCAC 83235  
Qy 41 ValGluAlaThyValAsnGlnIleIleValAlaLeuGluIleValGlyAlaSerIleValArg 60  
Db 83236 GTGGAAAGCAGCGGTTATCATGATCAAGCAAGCGCTGGAGCGGGTGGCCAGATATATGTCGT 83295  
Qy 61 ValSerValProThrMetAspAlaIleGluAlaPheLeuSerIleLeuGlnGlnValAsn 80  
Db 83296 GTTTCCTGACCAACATGACAGCCCGCCGAAAGCTTCAAGCTTATCAACAGCGGTTAGC 83355  
Qy 81 ValProLeuValAlaAspIleHisPheAspTyArgIleAlaLeuIleValAlaGluTy 100  
Db 83356 GTCCCGCTGGTCCGATATTCATTCCTTCGATCCGATCCGATCCGATCCGATCCGATCCGATCC 83415  
Qy 101 GlyValAspCysLeuArgIleLeuAsnProGlyAsnIleGlyAsnGluIleArgMet 120  
Db 83416 GGGGTGACCTGCTGGGTATTAACCCCGCAATATGCTAAACGAAAGCATTCGATC 83475  
Qy 121 ValValAspCysAlaArgAspIleAspIleProIleArgIleGlyValAsnAlaGlySer 140  
Db 83476 GTGGTGAATCGCGCGCGCAACAAATCCCTATCCGATCCGATCCGATCCGATCCGATCCGATCC 83535  
Qy 141 LeuGluIleAspLeuGlnGluIleTyArgIleGluProThrProGlnAlaLeuGlnGluSer 160  
Db 83536 CTGGAATAAGATCTGCAGGAAATAACGAGAGCCGACCGCGGCGCTGCTGATCA 83595  
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnHisLeuValSerVal 180  
Db 83596 GCATCGCGCATGTCATCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCTCATCT 83655  
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyArgLeuLeuAlaIleGlnIleAsp 200  
Db 83656 AAAGCTCGATATATCTCTTCCCTGATGATCTTATCTGTTGGCTAAACGATCGAT 83715  
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValIleSer 220  
Db 83716 CACCGCTGATCTGGGATCAACGAAAGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 83775  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
Db 83776 GCATCGGTTAGCTGCTGCTGTTATGGAAGGATCGGCAACCGCTGCGCTGATCGTTG 83835  
Qy 241 AlaAlaAspProValGluGluIleValGlyPheAspIleLeuIleSerLeuArgIle 260  
Db 83836 GCGCGGACCGGCTGGAAGAGATCAAGTCGTTGATTCCTGAAGTCCTGCTGCTGCTGCTGCTGCTGCT 83895  
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280  
Db 83896 CGCGCGCGCGGATCAATTCATCGCTGCGCGACCTGTTCTGCTCAGAGGTTGACGATT 83955  
Qy 281 IleGlyThrValAsnAlaLeuGlnGluIleValGluGluAspIleIleThrProMetAspVal 300  
Db 83956 ATCGGACGCTGAACCGCTGAGCGAGCGCTGGAAGATATCATCCCCGATGAGAGCTC 84015  
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 84016 TCATCATCGCTTGTGTGTGAACGAGCGCGGAGAGCGCTGATTCGACGCTGGGCGTG 84075  
Qy 321 ThrGlyIleAsnIleIleSerGlyLeuTyArgIleAspGlyValArgIleAspArgLeuAsp 340  
Db 84076 ACGGGCGGCAATTAAGAAAGCGGCTGTATGAAGAGCGGCTGCTGAAGGACAGATCGAT 84135  
Qy 341 AsnAsnAspMetIleAspGlnLeuGluIleArgIleArgAlaIleValAspSerIleLeuAsp 360  
Db 84136 AACGACGATATGATCGGACGCTGATGATCCGATTCGCGCGGAAAGTAAATGATCAACTTGAT 84195  
Qy 361 GluAlaArgArgIleAspValGlnGlnValGluIle 372  
Db 84196 GAAGCGGCTGATTCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 84231

RESULT 15  
AE016835 300431 bp DNA linear BCT 21-MAR-2003  
LOCUS Salmomella enterica subsp. enterica serovar Typhi Ty2, section 2 of  
DEFINITION 16 of the complete genome.  
ACCESSION AE016835 AE014613  
VERSION AE016835.1 GI:29136420  
KEYWORDS  
SOURCE  
ORGANISM Salmomella enterica subsp. enterica serovar Typhi Ty2  
Salmomella enterica subsp. enterica serovar Typhi Ty2  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Salmomella.  
REFERENCE  
AUTHORS Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,  
Burland, V., Kodoyanni, V., Schwartz, D.C. and Blattner, F.R.,  
Comparative Genomics of Salmomella enterica Serovar Typhi Strains  
Ty2 and CT18  
JOURNAL J. Bacteriol. 185 (7), 2330-2337 (2003)  
MEDLINE 22531367  
PUBMED 12644504  
2 (bases 1 to 300431)  
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,  
Burland, V., Kodoyanni, V., Schwartz, D.C. and Blattner, F.R.,  
Direct Submission  
Submitted (25-SEP-2002) Laboratory of Genetics, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
FEATURES  
source  
1. 300431  
/organism="Salmomella enterica subsp. enterica serovar  
Typhi Ty2"  
/mol\_type="genomic DNA"  
/strain="Ty2"  
/serovar="Typhi"  
/sub\_species="enterica"  
/db\_xref="taxon:209261"  
complement (67.. 486)  
/gene="yfiG"  
/locus\_tag="t0261"  
complement (67.. 486)  
/gene="yfiF"  
/locus\_tag="t0262"  
complement (67.. 1727)  
/note="corresponds to STY2842 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="thioredoxin 2"  
/protein\_id="AA067987.1"  
/db\_xref="GI:29136421"  
/translation="MNTVCTHCOAIRIPGDRLODAKCGRGHELFDGEVINATGET  
LDRLKDDLPVIVDFPAPMGPCGRNAPRIPEDEAERSGKRVKYNTERELRSARF  
GHSIPMIMFKHGQVVDMLNGAVPKAPFDSWLEAL"  
690.. 1727  
/gene="yfiF"  
/locus\_tag="t0262"  
690.. 1727  
/gene="yfiF"  
/locus\_tag="t0262"  
/note="corresponds to STY2841 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="putative RNA methyltransferase"  
/protein\_id="AA067988.1"  
/db\_xref="GI:29136422"  
/translation="MNDLKNKSGKVMYVRSDSDSKRTNPRTGKGGAPKAKRT  
DGRPARDERNNQSDRKHETSPMRTVSPAPGDEPEKVDHGSGSKSPIDEVDER  
ORAEPTRVGENAQQALFOSRPAIVRAFIQSVTRFKALEMMAANKAYHVVDA  
ELKASGTEHGGVCFILIKRNGITVKQWYKQADDCVLAEDVANPHNLGSMWSC  
AHGCVGVVQVADALLESQAIRTAEGGAHVPIIGESIVDLDDPFGAGYTVVTS  
SDGQALFSTTLPEKMLVLYGREYDYLPEAAREPDDLVCVINGTGVESLNVSVATGV  
LLEWVRONKA"  
complement (1843.. 2532)

CDS  
/gene="ung"  
/locus\_tag="t0263"  
complement (1843.. 2532)  
/gene="ung"  
/locus\_tag="t0263"  
/note="corresponds to STY2840 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="uracil-DNA glycosylase"  
/protein\_id="AA067989.1"  
/db\_xref="GI:29136423"  
/translation="MATELTWHDVLADEKQCFYITLTHTVAGERSQGTVPYPPQDV  
FNAFRTBLGVKVVILGQDPYHQAAGLAFVYRPGIAPPPLVMKELASIPG  
FYRPAHGYESNAROGVILLNVLTIVRAGASHASIGMETFDKVISLINOREGV  
FLMGSHACKGAIIIDPQHHLIKAPHPSPLSAHKGFPCNHPALTNQLEQHEKTI  
DNTPVLPARSE"  
2851.. 3234  
/gene="yfiD"  
/locus\_tag="t0264"  
2851.. 3234  
/gene="yfiD"  
/locus\_tag="t0264"  
/note="corresponds to STY2839 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA067990.1"  
/db\_xref="GI:29136424"  
/translation="MITGQITKANDDLNLSFWLLDSKGBARCTIVAKGPAEDBV  
AVSKGEIYREIPEMEKVEVGGQHLNVVLRRETTEDAVKPEKKPOLTRVSG  
VAVRPNLSLPBOQDVIAARTFRESL"  
complement (3296.. 3883)  
/gene="yfiK"  
/locus\_tag="t0265"  
complement (3296.. 3883)  
/gene="yfiK"  
/locus\_tag="t0265"  
/note="corresponds to STY2838 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="AA067991.1"  
/db\_xref="GI:29136425"  
/translation="MTPMLLSAFWTYTLITLTPGPNITLALSAATFANGFROSIRVLA  
GMSIGFLVYMLCAGIASLAVIDPAIHLISWGAAYILMLAWKIATSPADAEKVP  
KEVGFVWSFGQFVNVKILVIGITLSTFVLPODALNVIVGSILALIGTFGNVCW  
ALAGHLFOARFHHYGRQLNIIILALLVCAVARIFV"  
3986.. 4868  
/gene="yfiB"  
/locus\_tag="t0266"  
/note="corresponds to STY2837 from Accession AL513382:  
Salmomella typhi CT18"  
/pseudo  
complement (4901.. 6235)  
/gene="srnB"  
/locus\_tag="t0267"  
complement (4901.. 6235)  
/gene="srnB"  
/locus\_tag="t0267"  
/note="corresponds to STY2836 from Accession AL513382:  
Salmomella typhi CT18"  
/codon\_start=1  
/transl\_table=11  
/product="ATP-dependent RNA helicase SrmB"  
/protein\_id="AA067992.1"  
/db\_xref="GI:29136426"  
/translation="MTVTFSLELDESLLDALQDKGFTPTPAIQAAALPPALDGRDV  
LGSAPGTGCTAAVYLPALQHLDPFRKSGPRLIILPTRELAMQVADHARELAKH  
THLDIATITGVAVNMHAEVSENDIVVATGTGRLLOYIKENPFCRAVEETLILDEAD  
RMLDMGFADIDIHINGETRMKQTMLSATLEGDALKDPARELLDDPVEVSANPSTRE

gene

misc\_feature  
/note="Pfam match to entry PF00583 Acetyltransferase (GNAT) family, score 32.50, E-value 9.7e-06"  
complement (3790..3813)  
/gene="STY2723"  
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"  
complement (4420..5283)  
/gene="STY2724"  
complement (4420..5283)  
/gene="STY2724"  
complement (4420..5283)  
/note="Orthologue of E. coli P76563; Faeta hit to P76563 (287 aa), 92% identity in 286 aa overlap. Contains hydrophobic, possible membrane-spanning region"  
/codon\_start=1  
/transl\_table=1  
/product="putative membrane protein"  
/protein\_id="CAD02686.1"  
/db\_xref="GI:16503703"  
/db\_xref="SPTREMBL:O824S0"  
/translation="MRWQGRSNNVEDRRNRPGSPSLGGPRLPRGKGIILVVVAVAGYGVLDLIGLITOPVSQOSTSISPDDEAAKFTSVIATTEEDTGGI.PQKMGKGIQPKLVNRRKTRTGGCAGSWSGPFYCPADQTYITLSTYDVKKRLGADQFPAQGVIAEVEGHVQKLGIEPKVROLQONSQTEVNLSTRMELQADCFAGVGHSMQOQCVLEAGDLEALNMAOAIGDRLQOQGGRVVPSFTHTSEORYSWPKRGPDSGDPAQNTFGKNP"  
complement (5287..5292)  
/note="possible RBS"  
complement (5415..6128)  
/gene="STY2725"  
/note="synonym: purC"  
complement (5415..6128)  
/gene="STY2725"  
/note="Orthologue of E. coli purC (PUR7\_ECOLI); Faeta hit to PUR7\_ECOLI (237 aa), 93% identity in 237 aa overlap"  
/codon\_start=1  
/transl\_table=1  
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase"  
/protein\_id="CAD02687.1"  
/db\_xref="GI:16503704"  
/db\_xref="SPTREMBL:Q8XFE9"  
/translation="MQRQAEIYRGKAKTVVSTENPDLLVLEFRNDTSAGDGARIQEPDRKGVNKKFNHFWITKLAEGAIPTQMERLISDECLAKLEMPVECVNRRAGSLYKRLGVEEGMELNPIIFDLPLKNDALHDPMNNSYCTEPGVSONLARKMELYYKANDVAKLIPDAGLIVDPKLEPGLYKGEVVLDDPSPDSRLMDKETLDMKDXFRQSLGGIIEAYAVHRLGVKLD"  
complement (5421..6122)  
/gene="STY2725"  
/note="Pfam match to entry PF01259 SAICAR synt. SAICAR synthetase, score 414.20, E-value 1.2e-120"  
complement (5586..5612)  
/gene="STY2725"  
/note="PS01058 SAICAR synthetase signature 2"  
complement (5829..5873)  
/gene="STY2725"  
/note="PS01057 SAICAR synthetase signature 1"  
complement (6137..6141)  
/note="possible RBS"  
complement (6303..7337)  
/gene="STY2726"  
gene

RBS

misc\_feature

misc\_feature

misc\_feature

misc\_feature

gene

Alignment Scores:

Pred. No.:	2.9e-130	Length:	145050
Score:	1831.00	Matches:	365
Percent Similarity:	99.19%	Conservative:	4
Best Local Similarity:	98.12%	Mismatches:	3
Query Match:	98.12%	Indels:	0
DB:	1	Gaps:	0

US-09-921-992-50 (1-372) x AL627275 (1-145050)

Qy 1 Methisanginalaproileginalarglysertharargiletyvalaiglyasaval 20  
|||||

Db 64624 ATGCATACACAGGCTCCGATTCACAGTAGAATAATCAGCATATTCCTGGGAATGTC 64565  
Qy 21 |||||leGlyAspGlyValaProilealaValaGlnserMetThAsnthrarghtrAsp 40  
Db 64564 CCGATTGGGAGTGGCCCTTATCGCGTCAGTCATACATCAACGGGATCAGTGC 64505  
Qy 41 ValGluAlaThrValaAsnGlnIlelyAsalaLeuGluArgValaGlyValaAspIleValArg 60  
Db 64504 GTGAGACGACCGTTATCATGATCAAGCGCTGAGACGGCTTGCCACAGATATGTTCCG 64445  
Qy 61 ValSerValProTherMetAspAlaAlaGluAlaPheLyseuIlelyGlnGlnValaAsn 80  
Db 64444 GTTTCGTACCGCACATGACCGCCGGAAGCGTTTCAACCTTATCAACGACGAGTTACG 64385  
Qy 81 ValProleuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValaGluTyr 100  
Db 64384 GTCCCTCGTGGCCATATTCATCTTGACTACATCGGATTCGCTGAAGAGCGGAATAC 64325  
Qy 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
Db 64324 GCGGTGACTGCTTGCTGTTTAAACCCGCAATATCGTTAAGCAAGACGATTCGATG 64265  
Qy 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140  
Db 64264 GTGGTGACTGCGCGCGCAACAAACATCCATCCGATCGCGCTAAATGCCGATCG 64205  
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThProGlnAlaLeuGlnGlySer 160  
Db 64204 CTGGAAATGATCTGAGAAATATCGAGACCCGACCCGAGCGCTGCTGGAATCA 64145  
Qy 161 AlaMetArgHisValaAspHisLeuAspArgLeuAsnPheAspGlnPheLysValaSerVal 180  
Db 64144 GCGATCGCCATGTCATCATCTCGATCGCTCAACTTGTATGATGATCAAGAGACGTA 64085  
Qy 181 LysAlaSerAspValaPheLeuAlaValaGlySerTyrArgLeuAlaLysGlnIleAsp 200  
Db 64084 AAGCGCTCGATGTATTCCTTGCGCTTGATGTTTGTGCTTAACGATCGAT 64025  
Qy 201 GlnProLeuHisLeuGluLysIleThrGluAlaGlyValaArgSerGlyValaLysSer 220  
Db 64024 CACCCCTGATCTGGAGATCACCGAGCGGGCGGGGATGACGGGGGATTAAGTCG 63965  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValaSerLeu 240  
Db 63964 GCGATCGTTAGTGTCTGCTTATCGGAAGCATCGGACACGCTGCGCGATCGTTG 63905  
Qy 241 AlaAlaAspProValaGluGluIleLysValaGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 63904 GCGCGCGCGCGCATATTCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63845  
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
Db 63844 CGGCGCGCGCGCATATTCATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 63785  
Qy 281 IleGlyThrValaAsnAlaLeuGluGlnArgLeuGluAspIleIleThrPrometAspVal 300  
Db 63784 ATCGGACGCTGACCGCGCTGAGACAGCGCTGGAAGATATCATCCCGAGAGACGTC 63725  
Qy 301 SerIleIleGlyCysValaValaAsnGlyProGlyGluAlaLeuValaSerThrLeuGlyVal 320  
Db 63724 TCGATCATCGTGTGTGTGTGAACGGGCGGCGGAGCGCTGATTCGACGCTGCGCGCTG 63665  
Qy 321 ThrGlyGlyAsnLysSerGlyLysTyrGluAspGlyValaArgLysAspArgLeuAsp 340  
Db 63664 ACGGCGCGCAATTAAGAAAGCGCGCTGTATGAAGACGGCTGTGAAGACACATCGAT 63605  
Qy 341 AsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
Db 63604 AACGACGATATGATCCGACGCTTGAATCCGATTCGCGCGGAAGATGATCAACTTGA 63545  
Qy 361 GluAlaArgArgIleAspValaGlnGlnValaGluLys 372  
Db 63544 GAAGCGCTCGATTCGCTGCGAGGTTGAATAA 63509

Db	14529	GAAAGCGCTCGATTGACCTGCTGACGGTTGAAAAA	14494
RESULT 14			
LOCUS	AL627275/c		
DEFINITION	Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,		
ACCESSION	AL627275	145050 bp	DNA linear BCT 06-JUN-2002
VERSION	AL627275.1		
KEYWORDS	complete chromosome, segment 11/20.		
SOURCE	AL627275.1 GI:16503698		
ORGANISM	Salmonella enterica subsp. enterica serovar Typhi		
REFERENCE	Bacteria, Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
AUTHORS	1 (bases 1 to 145050) Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D., Wain, J., Churcher, C., Mungall, K.L., Bentley, S.D., Holden, M.T.G., Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Connor, P., Cronin, A., Davis, P., Davies, R.M., Dowd, L., White, N., Farrar, J., Feltham, T., Hamlin, N., Haque, A., Hien, T.T., Holroyd, S., Jagels, K., Krogli, A., Larsen, T.S., Leather, S., Moule, S., O'Garra, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrall, B.G.		
TITLE	Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18		
JOURNAL	Nature 413 (6858), 848-852 (2001)		
MEDLINE	21534947		
PUBMED	11677608		
REFERENCE	2 (bases 1 to 145050)		
AUTHORS	Parkhill, J.		
TITLE	Direct Submission		
JOURNAL	Submitted (23-OCT-2001) Submitted on behalf of the Salmonella sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT	E-mail: parkhill@sanger.ac.uk		
NOTES	Notes: Details of S. typhi sequencing at the Sanger Centre are available on the World Wide Web.		
FEATURES	(URL, <a href="http://www.sanger.ac.uk/Projects/S_typhi/">http://www.sanger.ac.uk/Projects/S_typhi/</a> ). Location/Qualifiers		
SOURCE	1. 145050 /organism="Salmonella enterica subsp. enterica serovar Typhi" /mol_type="genomic DNA" /strain="CT18" /db_xref="taxon:90370" 636..640 /note="possible RBS" 647..1003 /gene="STY2720" /note="synonym: yfjB" 647..1003 /gene="STY2720" /note="Orthologue of E. coli yfjB (YFPB_ECOLI); Faeta hit to yfjB_ECOLI (118 aa), 81% identity in 118 aa overlap" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="CAD02682.1" /db_xref="GI:16503699" /db_xref="SPTRMBL:O82483" /translation="MTLYGIKNCDTIKKARWDEHGIDYRFPHYVDGIDPLNTWPIALGQPLNTRGTGTRKRDDEHRSGITVDASAAALWGPRIKRPILLCARGKRWPLGSESRYOQFPEV" 995..1000 /gene="STY2720" /note="possible RBS" 1007..2134 /gene="STY2721" /note="synonym: dapE" 1007..2134		

*/gene="STY72721"*

*/note="orthologue of E. coli dape (DAPF\_ECOLI). Fasta hit to DAPF\_ECOLI (375 aa), 94% identity in 375 aa overlap"*

*/codon\_start=1*

*/transl\_table=1*

*/product="succinyl-di-aminopimelate desuccinyllase"*

*/protein\_id="CAD02683.1"*

*/db\_xref="GI:16503700"*

*/db\_xref="SPIRMBL:Q8Z4S2"*

*/translation="MSCVIELTQQILRRPSLDPDVGCALMIERIKIGFTIEHND  
FGDTDNFNFAWGREGTEIAFHGTVDVPAGDVAKINPPEPTIRIDGLFORGADVMG  
SLAMVAVEAFVQHAPHHRBRLAFLITSDSEASKGVTKVEALLMARRELDICLV  
MGDSSTEIVGVNVNKRGRSLTCNLTTIGHGVAYPHLDNPVRAARPLNELVALE  
WDRNEDFPAISMOIANIQATGSNNVIIPBELPQPMPFRSTELTDMEIERHALLE  
KHQLRYTDMWLSCGPFLTRKGVLDAVNVAIHENIKFKOLLTGCTSGDRFLRMNG  
AQVELGVNNTTIHKINECVNAADLOLLARYOKIMEJVA"*

*/misc\_feature*

*1031..1978*

*/gene="STY72721"*

*/note="Pfam match to entry PF01546 Peptidase M20,  
peptidase family M20/M25/M40, score 315.30, E-value  
7.4e-91"*

*1187..1216*

*/gene="STY72721"*

*/note="PS00758 Arge / dape / ACY1 / CPG2 / yscs family  
signature 1"*

*1295..1414*

*/gene="STY72721"*

*/note="PS00759 Arge / dape / ACY1 / CPG2 / yscs family  
signature 2"*

*2150..2153*

*/note="possible RBS"*

*2162..2362*

*/gene="STY72721"*

*2162..2362*

*/gene="STY72721"*

*/note="Similar to Escherichia coli similar to TR\_p7759  
(EMBL:D90875) (66 aa) fasta scores: E(): 1e-27, 95.5% id  
in 66 aa. Contains hydrophobic, possible membrane-spanning  
region"*

*/codon\_start=1*

*/transl\_table=1*

*/product="putative membrane protein"*

*/protein\_id="CAD02684.1"*

*/db\_xref="GI:16503701"*

*/db\_xref="SPTRMBL:O8XF02"*

*/translation="MDWLAKWMYLTVLPVGVLNVIXDKLRIDHKKFLANKRELPP  
HRFDNKDDDEDWDPKKDOPKK"*

*complement(2389..4404)*

*/gene="STY72723"*

*complement(2389..4404)*

*/gene="STY7273"*

*/note="Orthologue of E. coli YPF1 ECOLI; Fasta hit to  
YPF1\_ECOLI (671 aa), 72% identity in 662 aa overlap"*

*/codon\_start=1*

*/transl\_table=1*

*/product="conserved hypothetical protein"*

*/protein\_id="CAD02685.1"*

*/db\_xref="GI:16503702"*

*/db\_xref="SPTRMBL:O8Z4S1"*

*/translation="MSDIDLRALTSQMOTGITRIILVISGDAMGREAEAIRAVLP  
GDMMVAPARPQRCKTPQALQTLLRGREFRAITDANOGEPMAAFPALSLSLTQAGSWLL  
LLMPPVTWESEPRTDLSRWSDCAQRPTEPQFQNHKRTISRDPQTLMWRORPCWP  
SYPERERRPATGPDEPOEQAIIISRLPEMPGAVTAIAPGRSKSLAGCFIQMAGT  
AIVIPAAETADIILAAGERFCFMADALILASGARADMLVDEEAAILPPLLQSVS  
RFVPIILTTLTYEGEGTGRGLFKFCARFPOLHRFTLRBPVWAPECPLNISVALSI  
EDDAEFQAHPHGAIISAIFYQOMKGPALPRAYOLISGAHYRTSLDIRMWDAAG  
OHFIQATRNANNVAGALMVESGGLSAELSQAVCGRFFRRPNVAOSLASGSDPLLA  
TLVVRYSRIAAPAROREGIOQLACACEAOACGYVSFSGTYPEIMRVQRCGF  
LVVMGHSHRESGECYTAMALLPSSDGCLAOOEHRRLRDADITLWNNGEAIPLAA  
LLEBALUNDEDBRELVEGFAPAHLPLTSLGCHRLQYSALPLPALGRILEEKSADEL  
CARIRIGRKVIALAQAANAALINALDARTQRLADVWGCGDHAG"*

*complement(2878..3207)*

*/gene="STY2723"*

*\_misc\_feature*



```

/db xref="GI:16421061"
/translacion="MVRRIIRIVLVIVLVSVVAAGSADNSANDSHSSLPDLASRS
AKREBENKSKSREKAGDYFINSATOGFENLTPBEAESABYLAQINQISASQVLE
GMSPYKRTITSLSVEGDDLDSSSLDFIPWYDNOSTLFSQISAKRKDRITIGNG
LGVQVQVNNVLGLGNAFYDYDFRGRHRLGLGEAMVDYKTFSGNTHHPSDKWDED
FDEYEREPARGMDIRMESMLPEYQOLGAKLYEYQYDGEVALGTQNDLQCPDPAKED
LEAYPLVLTIVGTDYKAGTSDNSDFSNATVANYQIGPLAQLDPEWVKIOHLSMR
TDFVDRNPFILLEYREKQPLDVTLMKADATNHEPECVIDTPEEAALGEKCKMTYNA
LIMHXYKISASMOAKNNARATVMPVXKNAALTEGNSNWNVLPAWVAADPEORT
ALNTWKRTMRLEDEKKNQNSQVEITVQODRKIELVDNIAIDTDSHSHESALAD
GEGVAVDILLITDSFGDSTDNNGNELVDAMTPVLVDSNKKVTLAQPTCTTPTCVF
IASRDEKAGVTLLSTLPGTFRMKAKADAGDSYVDVPTIGNLSALNVAIOVNA
NPVNLGKEDKHPTVNNARYFLLMRDKNKGVSQMSQLEEMALVDYQWMEFTGOST
NGHTGALANTNEDVLVPTNKEAOKFAANVEDGVGYSIRVYSQK"
complement (9037..9042)
/gene="sinh"
/note="putative RBS for sinh; RegulonDB:STWS1H002624"
complement (9382..9615)
/gene="yfgJ"
/note="synonym: STM2518"
complement (9382..9603)
/gene="yfgJ"
/note="similar to E. coli orf, hypothetical protein
(AAC75563.1); Blastp hit to AAC75563.1 (83 aa), 63%
identity in aa 13 - 83"
/codon_start=1
/transl_table=1
/product="putative cytoplasmic protein"
/protein_id="AAL21412.1"
/db xref="GI:16421062"
/translacion="MEITCPVCHHALERNQDTAHCTCAKDFSLQALCPDQPLQVL
KACGAVDFPCONGHGLISKXNVNFISSDQ"
complement (9610..9615)
/gene="yfgJ"
/note="putative RBS for yfgJ; RegulonDB:STWS1H002625"
complement (9692..11176)
/gene="engA"
/note="synonym: STM2519"
complement (9692..11164)
/gene="engA"
/note="similar to E. coli putative GTP-binding factor
(AAC75564.1); Blastp hit to AAC75564.1 (503 aa), 97%
identity in aa 14 - 503"
/codon_start=1
/transl_table=1
/product="putative GTP-binding protein"
/protein_id="AAL21413.1"
/db xref="GI:16421063"
/translacion="MVPVVALVGRPNVSGKSTLFNRRLTRTRDALVADPPGLTRDRKGR
AYVGRFRFICIDIGIDGTEDGVTWMAEDSILAIERADVLVLMVARAGIAMDADI
AKHLRSREKPTPLVANKTDGLDPQAVVDYSGIGITVPIAASHRGVLSLLEHVL
PMWDVAPOEVEDDADAYMAOFEBQGESEAPEDDPQSLPKLALVGRPNVSGKSTL
TNRLGEBRVVVDMPGTRDSIYIPMERDEREYVLIDTGVRRKGITDAVKSFTV
KTIQALIDANVVLVIDAREGISDQSLGFIINSRSLIVVNKKDGLSOVEKEV
KETIDPRLGIDFARVHFISALHSGVGNLFESVREAYDSTSRVSTAMLTRIMTAV
EDHOPPLVGRGRVTKYAHAGVNPPIVHGNQVOLDPSYKRYLMNVRKSLLEWVG
TPRIQPKESBNPRANKRNTLTPQMKRKLKMKIKSK"
complement (11171..11176)
/gene="engA"
/note="putative RBS for engA; RegulonDB:STWS1H002626"
complement (11283..12471)
/gene="yfgL"
/note="synonym: STM2520"

```

## Alignment Scores:

```

Pred. No.: 5.3e-132 Length: 23647
Score: 1840.00 Matches: 367
Percent Similarity: 99.46% Conservative: 3
Best Local Similarity: 98.66% Mismatches: 2
Query Match: 98.61% Indels: 0
DB: 1 Gaps: 0

```

US-09-921-992-50 (1-372) x AE008814 (1-23647)

```

QY 1 MetHisAsnGluAlaProIleGlnArgLysSerThrArgIleTyrValGluAsnVal 20
Db 15609 ATCATTAACACGAGCTCGATTCAACGTAGAAATTCACACGTATTACCTGGGAATGTC 15550
QY 21 ProIleGlyAspGlyValAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 15549 CCGATTGGGAGATGTCGCCCTATCGCGTCGATGATGATCAACGCGTACCATGAC 15490
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluLysValGlyAlaAspIleValArg 60
Db 15489 GTGGAGGCACCGTTATATAGATCAAGCAAGCGCGAGACCGGTTGGCCAGATATGTCGT 15430
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
Db 15429 GTTTCCTACCAATAGACGCCGCCGAGAGCTTACAGCTTACAAACGACGCTTAC 15370
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 15369 GTCCCGCTGGTGGCGATATTCATCTTCACTACCGCATTGGCGTGAAGTAGCGAATAC 15310
QY 101 GlyValAlaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluLysIleArgMet 120
Db 15309 GCGGTGACTGCTGCGTATCAACCCGCAATATCGTAAACGAAACGTATTCGTATG 15250
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 15249 GTGTGTGACTGCGCGCGCAAAACATCTTATTCGATTCGCGCTTAACGCCGATTCG 15190
QY 141 LeuGluLysAspLeuGlnLysTyrGlyGluProThrProGlnAlaLeuLeuGlnSer 160
Db 15189 CTGAAATAAGATCTGACGAAATAATACGGTAGACGACGCGGACGCGTGTGTAATCA 15130
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnHisLysValSerVal 180
Db 15129 GGCATCGCGATGTCATCATCATCGATCGATCGATCACTTCAATGATCAAGTAAAGACGTA 15070
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 15069 AAGCGTCCGATGATATTCCTTCCGCTTGAGTCTTATTCGTTTGTGGCTAAACGATCAT 15010
QY 201 GlnProLeuHisLeuGlnLysIleThrGluAlaGlyValAlaArgSerGlyValAlaLysSer 220
Db 15009 CACCGCTGATCTGGGATCATCCGAAAGCGGGCGCGGTACCGGGCGGTCAAGTGC 14950
QY 221 AlaIleGlyLeuGlnLysLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 14949 GCGATCGGTTTGGTCTGCTGTTATCGGAAGGATCGGAGACACGCTGCGCGTATCGTTG 14890
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 14889 GCGGCGGACCGGTGGAAGAGATCAAAAGTTCGATTGATTTTGAAAGTCCGCGCATTT 14830
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280
Db 14829 CCGCGCGCGGATCAATTTCATCGCTCGCGACCTGTTCTCGCAGAGGTTGACGATT 14770
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 14769 ATCGGACCGGTGAACCGCTGCGCGCGCTCGGAAGATATCATCAACCGAGAGACGTC 14710
QY 301 SerIleIleGlyCysValValAlaAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 14709 TCGATCATCGGTTGCTGATGAACGCGGCGGCGGAGCCGCTGATATCGACGCTGGCGGTG 14650
QY 321 ThrGlyAlaAsnLysLysSerGlyLeuTyrGluAspGlyValAlaArgLysAspArgLeuAsp 340
Db 14649 ACGGCGGCAATATGAAGAAAGCGGCTGTATGAAGACGCGTGGTGAAGACAGACTCGAT 14590
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaValAlaSerGlnLeuAsp 360
Db 14589 AACGACGATATGATCGCGACGCTTGAATCCCGTATTCGCGGAAACAGTCACTTGAT 14530

```

```

Db      189 AACACGATTCAGTGCACCCTGGAAACACCGACTTCGTCCGAAGAACCAGTCAGCTGAC   130
Qy      361 GUAAlarGaRgllEapVaIGnclnvaJgtulyS 372
LOCUS   129 GAACGCCGTCAATTGACGTTCAGACAGGTGGAAA 94
RESULT 13
AE008814/c AE008814          23647 bp     DNA        linear    BCT 23-APR-2003
DEFINITION Salmoneella typhimurium LT2, section 118 of 220 of the complete genome.
ACCESSION AE008814 AE006468
VERSION   AE008814.1 GI:16421058
KEYWORDS
SOURCE    Salmoneella typhimurium LT2
           Salmoneella typhimurium LT2
           Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
           Enterobacteriaceae; Salmoneella.
REFERENCE 1 (bases 1 to 23647)
AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
           Latreille,P., Courtrey,L., Potwollik,S., Ali,U., Dante,M., Du,F.,
           Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
           Grewal,N., Milvanev,E., Ryan,B., Sun,H., Flores,L., Miller,W.,
           Stonelung,T., Nhan,M., Waterston,R. and Wilson,R.K.
TITLE      Complete genome sequence of Salmoneella enterica serovar Typhimurium
           LT2
JOURNAL   Nature 413 (6858), 852-856 (2001)
MEDLINE   21534948
PUBMED    11677609
REFERENCE 2 (bases 1 to 23647)
AUTHORS   The Salmoneella typhimurium Genome Sequencing Project
TITLE      Direct Submission
COMMENT    Submitted (29-MAR-2001) Genome Sequencing Center, Department of
           Genetics, Washington University School of Medicine, 444 Forest
           Park Boulevard, St. Louis, MO 63108 USA
COMMENTS  Supported by NIH grant SU 01 AI43283

COMMENT    Coding sequences below are predicted from manually evaluated
           computer analysis, using similarity information and the programs;
           GLIMMER: http://www.tigr.org/software/glimmer/glimmer.html and
           GeneMark; http://opal.biology.gatech.edu/GeneMark/

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
and Pedro Romero and Peter Karp at EcoCyc;
http://ecocyc.PangeaSyste.ms.com/ecocyc/

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
RegunoldBj;
http://klinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

Location/Qualifiers
1..23647
/oranism="Salmoneella typhimurium LT2"
/mol_type="genomic DNA"
/stain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:9928"
/note="LT2"
/complement(110..5711)
/gene="rfaA"
/note="synonym: STM2515"
complement(110..5707)
/gene="rfaA"
```

/note="RatA [gi|4583529]"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative outer membrane protein"  
 /protein\_id="AAL21409.1"  
 /db\_xref="GI:16421059"  
 /translation="MDRRROGVKVVHFLTLTLIMFAVISLAPQALLKGGTMOELNSVTL  
 GAVNCTAPLALGALLPLYQSGTLLDPESKTHDIESAMPDPDSADATSTSKRAVNSTYTT  
 EGDSESDPPTLAMERQEPVAGLVWADATPTDPLSPQVFNLTFCQNLAGRLVAM  
 AQVEDETENMLRQPTGTGVNVAITPLLSKVALNKPATIDPVTYVQGHVAGLVAM  
 SKVQKESITLTIITKACDGTPEGNAEPFVIRKQAEKRGCVNANFVRVGDTELT  
 QTEYGVGDANAGNAVTVVTOKEQEGVTPLVASATACGVGVPOSSALVLSHAGGMOEPI  
 ASMGMHESATITLNGITETPTFRPKLAASGADKSDVDTNETMALFTMSGANDHD  
 ILPDAEQVQKREHSTLETTTGMVAGDAEYASTKQSLSGTHAAVHMASVYRRA  
 DSDTLVSCVQKRAHPHPIQTLTSPQYKAQVESIDLTMTVVDRTQKPLPRVWE  
 LEPIDAKRKEQEDAMDNDNLVTVDSBGMSSPEHYTGVDVNGAHLTLKNSGGLG  
 VETPRILAMPDEGDNELSPSVLFTVVTSPDVGAMMGHMRGVDAQNLKRPPLLA  
 VEAHSGEENNEEWATPNSVASATACGVGVPOSSALVLSHAGGMOEPI  
 MPTBEYPIAASDSDSGYHVNLKENGDSGKFPDPNTYITCSANEMVAALDYFNDP  
 KNADKVEQYKRNMIHSHNALNGMAGICGNTDFTTWANGRRDGLTFTFTTSCNEQF  
 DQVGVAAQVYHGTIDANGATIIILTKQKGVGLTPLNLAIVDSLISTPVSRSKVF  
 ATSPPTPAKMGHMDITVGDWTFERPKLAGEVSNPLTQDSNETMARVASHDVAV  
 GNPDGCAANRLPRIDQLBALYNASGGKSLISOGMTVNSVYSOGATTKMLJA  
 LTNCEFSFANVSIYASCLASDNVVASITIEPNPNOWDGSVYHAKKKEBTML  
 KYTVKDVSGNIPPAFPVLTREGDGYDRGEKTYKQDDDLQIGITPVTVDSLANMT  
 TKMSQSTIDTGRITLITVRPDTHGRITAINNLITENAVASIDTITVTVSPVSA  
 RMWGMASSLTADGAVAYORPLVLAELSDNTASKQETNEVAFPGALDEGNAPR  
 CAAGYPAVEALDLSKYPSERTINACQNPVYVYSGNSNSTSFSSGALDEYVAD  
 LADGSRSENEASTTAMOVQICATTPLOATQITLTSPOAMDDAIOVKAKNSESIPL  
 LITTTGADGNTVAFTSLIKRDAKARNPDNKNFVANGNMVYTLPTGAQOOFYAT  
 SVLTGATGADDTLTLAERBGILSKNOLTRANIDTPTAISLSPVFTVYLTSPSDKA  
 NMYGMEPTFPAASGAEPKPLVAGESSSEAHNTDITYETENMETIMVNSFNTGNGCP  
 NMQMAALIDFTFALYNHDSRGVATIDIGLPVGRWAGDSLKSGSLTKQYKDLTKGN  
 YMSNENPNVYLOCLTTHSPSGVATIDIALSSPAMNADKRAAAKKEITLPMVYTVYNDG  
 QPQKGLVAILTRDYAVSRGAVDKQYIEPGYGEVPEPTTSPAMMLLPAVPAQAVAF  
 NONGSLSTKMSGFGTGDGKLFRTLTPDKSKLGKTPVSAALANQDEASADALFTVOT  
 SPDIPTYASVYMGHMDTVQVNGVTILRRPYLAELLSAARDPVPENNEPFWNGYVYQSH  
 VETSITHLICSGENIASSLDDIKALQSVITGLVQNPFTTSMWYVSODEQSNKYICSENE  
 VTGQTTCTREKRAVTSIGSGSCRP"  
 complement (5706..5711)  
 /gene="RatA"  
 /note="putative RBS for ratA; RegulonDB:STWS1H002622"  
 complement (5827..6793)  
 /gene="sin1"  
 /note="synonym: STM2516"  
 complement (5827..6786)  
 /gene="sin1"  
 /note="sin1 [gi|4583530]"  
 /codon\_start=1  
 /transl\_table=1  
 /product="putative outer membrane protein"  
 /protein\_id="AAL21410.1"  
 /db\_xref="GI:16421060"  
 /translation="WQATVVRRLTKVALALVNGVCAAPVANGNLSKSGOMQIVSQ  
 TGIQGVAVITRAPADTKADTDKDHVTTITDREGKRLVTGDKQFHVGDVYVYMAIG  
 DTEDDLDTDNWATKATVAKWSFSDQNSDPRKDLGTGSYSEIQAADARVYIGIKITPTT  
 TTGPDAVATIELKDLSTDGAGSGSDDEIEPEGVVENHVVHLEKDSNNLKNKST  
 TLKNTITQVILLKDKNGNGFYDAGENVQDYVRKMFVTSYKLAGTGTGIVNESVN  
 DKDLIVPTNNAEATFEGADGVTVSDDGVGSGLSIDYKRC"  
 complement (6788..6793)  
 /gene="sin1"  
 /note="putative RBS for sin1; RegulonDB:STWS1H002623"  
 complement (6844..9042)  
 /gene="sinH"  
 /note="synonym: STM2517"  
 complement (6844..9036)  
 /gene="sinH"  
 /note="similar to Escherichia coli inticin and Yersinia  
 pestis invasins proteins; [gi|4583531]"  
 /codon\_start=1  
 /transl\_table=1  
 /product="sinH"  
 /protein\_id="AAL21411.1"

/transl\_table=11  
/product\_id="ISSf14orf"  
/protein\_id="AAPI7892.1"  
/db\_xref="GI:30042167"  
/translation="MTESDSYVESVQVFIGVDVGDKTHHVAIVNSGRKLPFDKALPNDE  
NKRRLSIDKQKHQIILVDDOPATIGALVAVARSQVAVGLPGIAMRRIIDLHAG  
EATDVEDAAIIEAARTLPHALRTILAEQIAELSMCGPDDILAQCTQASNRIR  
GLATQIHPAPERIEVPRLEHVAVIDLQRYSPSEPKLASGEKKLAOLCLAPRLGKR  
LAADIQALAEQTVVVGNTAAAVLPRALQILTLRKQPDVALEQVAVLHNPY  
VUTSMGVGVRTAARLITTEVACRAFASAAHLAAYAGIAPTRRSGSIRGEHNRGN  
KAKRAFLFSAFALRDLPSRAYYTRKMSQGRNDLALARRCDVLFAMWRDGF  
YTPQGS"  
7960..8442  
/gene="ssea"  
/locus\_tag="S2740"  
/note="truncated"  
/pseudo  
complement (9261..10037)  
/gene="sseb"  
/locus\_tag="S2741"  
complement (9261..10037)  
/gene="sseb"  
/locus\_tag="S2741"  
/function="phenotype; Not classified"  
/note="residues 1 to 258 of 258 are 94.96 pct identical to  
residues 4 to 261 of 261 from Escherichia coli K-12 :  
B2522"  
/codon\_start=1  
/transl\_table=11  
/product="enhancer of serine sensitivity"  
/protein\_id="AAPI7893.1"  
/db\_xref="GI:30042168"  
/translation="MSETKLELDLEKATEPAPAPRTILESTVWVPGTAQGE  
AVEDSALDQHEKEDGTSVITPFTSLLELQAVEDQAFVMPRTTLEMTLGETL  
FLNAKLPTEKEMPRETSLTIGENPLSQETLEGGESLISVAEPDQAMDSLT  
LFTTIRKVKAFICSIKENEAQPNLIGLEADGIEIILQATGSVATDTLPDEBID  
ICQKKGKGISHFITEHIAFYERERGGFLRDFQNRIT"  
10137..10170  
/note="RBP (repetitive extragenic palindromic) element,  
contains 1 RBP sequence"  
complement (10179..11462)  
/gene="pepb"  
/locus\_tag="S2742"  
complement (10179..11462)  
/gene="pepb"  
/locus\_tag="S2742"  
/function="putative enzyme; Not classified"  
/note="residues 1 to 427 of 427 are 92.50 pct identical to  
residues 30 to 456 of 456 from Escherichia coli K-12 :  
B2523"  
/codon\_start=1  
/transl\_table=11  
/product="putative peptidase"  
/protein\_id="AAPI7894.1"  
/db\_xref="GI:30042169"  
/translation="MTBAMKITLSTOPADARWGEKATYSINNDGITLHNGDIGLI  
ORARKIDLGITGVQLSGGMDADRCMAWQCYKPKGRKTEMEDLDQAOEIDN  
RLMTIDVDRITNAPBELSPQLAQAVDLISVAGDRVTRITKEBDRREGYKGL  
HTVGRGERSPVLLADYNPTGDEKAVAVLGVGKGTTPDSGGYSIKQTAFMDSKSD  
MGGAATVGTALAFIMRGLNKRVLFLCCADNLISGNAFLGDIITYRNKKEVWNT  
DAGRLVADGLIDASOKPKEMIIDATLQAKTALGNDYHALFSDDDLARLLAS  
ASQENEPFMRLLPABFHSQLPSPNPAELNNTGSAAYAGASTAAGFLSHVENYQCGM  
LHIDCATYRKAVPEQMSAGATGLGRTIANLTA"  
complement (11521..11721)  
/gene="yfhj"  
/locus\_tag="S2743"  
/gene="yfhj"  
/locus\_tag="S2743"

Alignment Scores:  
Pred. No.: 1e-131 Length: 290380  
Percent: 1855.00 Matches: 370  
Score: 99.73% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 1  
Query Match: 99.41% Indels: 0

DB: 1 Gaps: 0  
US-09-921-992-50 (1-372) x AE016987 (1-290380)  
Qy 1 MethiAsnGlnAlaProIleGlnArgLysSerThrArgLLeTyValIAsnVal 20  
Db 1209 ATGCATTAACAGCGCTCCATTCAACGTAGAAATCAACAGTATTACGTGGGAATGTC 1150  
Qy 21 ProIleGlyAsnGlyValaProIleAlaValaGlnSerMetThrAsnThrArgTrpTrp 40  
Db 1149 CCGATTGGCGATGGTCTCCATCGCGCTGATCATCAATACCAATACCGGTAGACAGAT 1090  
Qy 41 ValGluAlaThrValaAsnGlnIleLysAlaLeuGluArgValaGlyIAspIleValArg 60  
Db 1089 GTCCAGCAACGCTCATCAATCAAGCGCTGGAACGGCTTGCCCTGATATCGTCCGT 1030  
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuLLeLysGlnGlnValaAsn 80  
Db 1029 GTCTCCGTACCGACGATGACCGCGGAGAACGCTTCAAACTCATCAAAACGCGGTTAAc 970  
Qy 81 ValProLeuValaIAspIleHisPheAspTyArgIleAlaLeuLysValaIAspGlyTy 100  
Db 969 GTCCCGCTGCTGCTGACATCACTTCACTGATGCGATTGCTGAAGTAGCGGAATAC 910  
Qy 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnGluArgIleArgMet 120  
Db 909 GCGGTGATGCTGCTGCTTAACTTGAATATCGTATCAAGAGCGTATTCCTCATG 850  
Qy 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140  
Db 849 GTGGTGAATGCTGCGCGCATAAACATTCGATCGTATTGGCTTAAACGCGGATCG 790  
Qy 141 LeuGlnLysAspLeuGlnGlnLysTyArgLysValProThrProGlnAlaLeuLeuGlnSer 160  
Db 789 CTGGAATAAGATCTGCAATAAAGATAGCGAACCCAGCCGCGGCTGCTGGATCC 730  
Qy 161 AlaMetArgHisValaAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
Db 729 GCATCGCTCATGTTGATCATCTCGATCCGCTGAATTCGATCAGTTCAAGTCACCGTG 670  
Qy 181 LysAlaSerAspValPheLeuAlaValGlnSerTyArgLeuLeuAlaLysGlnIleAsp 200  
Db 669 AAAGCGCTGACGCTCTCCCTCGCTGTGAGTCTTATCGTTGCTGCAAAACAGATCAT 610  
Qy 201 GlnProLeuHisLeuGlnLysLeuThrGluAlaGlyValaArgSerGlyAlaValLysSer 220  
Db 609 CAGCCCTCATCTTGGGATCACCGAAGCCGGTGTGACGACGCGGCGCAGTAAATCC 550  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeu 240  
Db 549 GCCATTGTTAGTGTCTGCTGCTGCTGCGAAGCATCGGACACGCTGCTGATCGCTG 490  
Qy 241 AlaAlaAspProValGlnGlnLysValaGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 489 GCGCGCATCCGCTGCAAGATCAAGATCGGTTGATTTTGAATTCGCTGCGCTATC 430  
Qy 261 ArgSerArgGlyIleAsnMetIleAlaCysProThrCysSerArgGlnGlnPheAspVal 280  
Db 429 CGTTCGAGGAGTAACTTCACTTCGCTGCTGCGAAGCATCGGACACGCTGCTGATCGCTG 370  
Qy 281 IleGlyThrValaAsnAlaLeuGlnGlnArgLeuGlnAspIleIleThrProMetCAspVal 300  
Db 369 ATCGGATCGGTTAACCGCTGAGCAACCGCTGGAAGATATCATCTCCGATGACGCTT 310  
Qy 301 SerIleIleGlyCysValaValaAsnGlyProGlyGlnAlaLeuValaSerThrLeuGlyVal 320  
Db 309 TCGATTATCGGCTGCGCTGTAATGAGCCAGGAGAGCCGCTGTTCTACACTCGGCGTTC 250  
Qy 321 ThrGlyGlyAsnLysSerGlyLeuTyArgLysGlyValaArgLysAspArgLeuAsp 340  
Db 249 ACCGGGGGCAACGAAAGCGGCTCTTATGAAGATGGGCTGGCAAAACCGCTGCGAC 190  
Qy 341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360

Db	Accession	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	PMID	FEATURES	source
Qy	6662	ACC6GCGGCAACAGAAAGAGGGCGCTCTATGAAAGATGGCGGCGCAAGACCGTCTGGAC	6623												
Qy	341	Aenhbapbwtcttleapbglhucgluaalarglileatgalaalyalserglinleuasp	360												
Db	6622	AACAACGATATGATGCGACCGACTGTGGAACAGCGCATTCGTGGAAAGCCAGTACGCTGGAC	6563												
Qy	361	GIUuLaARgArGlleApSpAlGlnGlnAlaGluLyS	372												
Db	6562	GAAAGCGCTCGAATTGACCTTCAGCGAGGTGGAAAA	6527												
RESULT 12															
AE016987/c															
LOCUS	AE016987	290380 bp	DNA	linear	BC1	22-APR-2003									
DEFINITION	Shigella flexneri 2a str. 2457T	section 10 of 16 of the complete genome.													
ACCESSION	AE016987	AE014073													
VERSION	AE016987.1	GI:30042162													
KEYWORDS															
SOURCE	Shigella flexneri 2a str. 2457T														
ORGANISM	Shigella flexneri 2a str. 2457T														
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.														
AUTHORS	1 (bases 1 to 290380)														
	Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G. II, Rose, D. J., Darling, A., Mau, B., Pern, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C. and Blattner, F. R.														
	Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T														
	Infect. Immun. 71 (5), 2775-2786 (2003)														
	2 (bases 1 to 290380)														
	Wei, J., Goldberg, M. B., Burland, V., Venkatesan, M. M., Deng, W., Fournier, G., Mayhew, G. F., Plunkett, G. II, Rose, D. J., Darling, A., Mau, B., Pern, N. T., Payne, S. M., Runyen-Janecky, L. J., Zhou, S., Schwartz, D. C. and Blattner, F. R.														
	Direct Submission														
	Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA														
	Location/Qualifiers														
	1. 290380														
	/organism="Shigella flexneri 2a str. 2457T"														
	/mol_type="genomic DNA"														
	/strain="2457T"														
	/serotype="2a"														
	/db_xref="caxon:198215"														
	complement (82. 1209)														
	/gene="gcpB"														
	/locus_tag="S2733"														
	complement (82. 1209)														
	/gene="gcpB"														
	/locus_tag="S2733"														
	/note="residues 1 to 372 of 375 are 96.23 pct identical to residues 1 to 372 of 372 from Escherichia coli K-12 : B2515"														
	/codon_start=1														
	/transl_table=1														
	/product="hypothetical protein"														
	/protein_id="AAP17888.1"														
	/db_xref="GI:30042163"														
	/translation="MHNQAPLQRRKSTRIVGVNPIQDGADIAVQSMTRTTTDEATVNOIKALERVADIVRSVPTMDAAEAFKIKQEVNVPVADIHFDV														

```

/function="putative membrane; Not classified"
/note="residues 1 to 337 of 337 are 80.71 pct identical to
residues 1 to 337 of 337 from Escherichia coli K-12 :
B2516"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAP17889.1"
/db_xref="GI:30042164"
/translation="MNTETATHDQNEALTTGARLRNAREQLGSOQAVERLCLKVSTV
RDIEDKRPADPLASTFLRGLYRSYARLVH1PEBELPLGLRKAQPLRAKAPAPMOSFEL
GKRKKRQKQKPTFMVLVFLVIGLSGAMMMDHKAQOEITTMADQSSAE1LSNNSQ
GOSVLYNTSTTDATSTSTPASVDGTTATNTQPAVAPAPAYDQPMNAVVSQNAV
DTA1LPVTAA1TTPDGAAPLPFTDQAGVTTTPADPNALVMPFADCHLEVDTAGKLF
SGMQKDKDNLNLTGQAPFKLIGAPAAVQIYQKQPVDSFRFTRTNQARLTNAEOS
PAQ"
/complement(2534. .3688)
/gene="yfgB"
/locus_tag="S2735"
/complement(2534. .3688)
/gene="yfgB"
/locus_tag="S2735"
/note="residues 1 to 384 of 384 are 99.47 pct identical to
residues 1 to 384 of 384 from Escherichia coli K-12 :
B2517"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAP17890.1"
/db_xref="GI:30042165"
/translation="MSRQLYTPRENVTTKDKKINLIDNRQMRPEFKDGEKTFPADO
VMKMYTHYCNFDEMTDINKVLRGKLKEVAELTAPRVEEQSSDGTIKMATAVEDQ
RVEYVYIEDDRATLCVSSQVGCALCEKFCSTAQGNRLRVSEIIIGQWRAKTGV
AAKYQGP1P1VNNVMMGEBELNNVNVAMEIMLDDFGGLSKRVTLTSGVPA
LDKLGKIDVLA1SLHAPNDEIRDEIVPINKYNIETFLAAYRYLEKSNANQGRVT
IEVWLMDHVNQTHAHQIAELIKDTPCKINL1PMNFPAPAGRSSNSR1IDRFSKVL
MSYGTITVTRKTRDDDD1DAAGQLAGDV1DRTKRTLAKRNGEAIIDKAY"
/complement(3838. .4269)
/gene="ndk"
/locus_tag="S2736"
/complement(3838. .4269)
/gene="ndk"
/locus_tag="S2736"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="residues 1 to 143 of 143 are 100.00 pct identical
to residues 1 to 143 of 143 from Escherichia coli K-12 :
B2518"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAP17891.1"
/db_xref="GI:30042166"
/translation="MATERPFSIIKPNAAKNVIGNTFAFPEAGFIVTGKMLH1TV
EQARFVHEHKGKPFDDGLVEFMSTGPIVIVSLEGENAYVRHDLGATNPAAALACT
LRADYADSLTENGTHGSDSVESAARE1ATVFFGEVCPRT"
/complement(4418. .6517)
/gene="bpbC"
/locus_tag="S2737"
/note="truncated"
/pseudo
6514. .7961
/note="insertion element"
/insertion_seq="ISSF14"
/complement(6534. .7730)
/locus_tag="S2738"
/complement(6534. .7730)
/locus_tag="S2738"
/function="TS, phage, Tn; Transposon-related functions"
/note="residues 1 to 398 of 398 are 90.95 pct identical to
residues 1 to 398 of 398 from GenPept :
>gb|AAK18438.1|AF348706.127 (AF348706) putative
transposase for IS110 [Shigella flexneri]"
/codon_start=1

```

```
/product="orf, conserved hypothetical protein"
/protein_id="AAM4059.1"
/db_xref="GI:24052933"
/translation="MEIYENENDQVEAVKFPFANGKALAVGGLIGVALTGRWYNS
HVDVASASLAVQNAVTAVSEKPDSPFAEKFAENKNTYALASLELAQDFVKN
ELKKAQAQOQGLADTSDENLKAVINLRARVOVLKQADAPLKTIDTITIGREMAAIV
ADRGELALSKGKQKARSAMEAGKSDVTPALSEMMQMKINMLSI"
complement(5138..6412)
/gene="hies"
/locus_tag="SF2560"
complement(5138..6412)
/gene="hies"
/locus_tag="SF2560"
/locus_tag="SF2560"
/notes="Residues 1 to 424 of 424 are 99 pct identical to
residues 1 to 424 of a 424 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289067.1"
/codon_start=1
/translation="hisidine tRNA synthetase"
/product="hisidine tRNA synthetase"
/protein_id="AAM4060.1"
/db_xref="GI:24052934"
/translation="MAKNTOAIRGMDYLFGEITAIWIRIGITLKNVLGSGYSEILP
IVEQTLPEKAIQGEVDVVEKEMYTEEDRGDSLTLRPEGTACVRAGLEHLLYQOE
QRLMYIGPMFRHERPQKGRYRQPHOEGEVGQSPDIDAEILMLTARMRALGIEH
VTELENSIGLEKARANYRDAVALEQHEKLEDECKRMYTNPRLVLDSEKNEVAL
LNDAPALGDLDESRHEFHAGLCKLESAGIATVNORIRGIDYNNRTPEVMTSL
GSGCTVAGGRYDGLVYEQLGSRATPANGFAMGLPRYLIVQANPERKADPVYDIYV
ASGADTOSAMALAEIRLDELGVKLTNNGGKFNKQFARADKMGARAVVIGSESV
ANGTAVVKDLRSEQITVAQDSVAHLRLTLG"
complement(6515..7642)
/gene="gcpB"
/locus_tag="SF2561"
complement(6515..7642)
/gene="gcpB"
/locus_tag="SF2561"
/notes="Residues 1 to 372 of 375 are 99 pct identical to
residues 1 to 372 of a 372 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289068.1"
/codon_start=1
/translation="orf, conserved hypothetical protein"
/product="orf, conserved hypothetical protein"
/protein_id="AAM4061.1"
/db_xref="GI:24052935"
/translation="MHNQAPLORRKSRIIVGNVPIGDAPIAVOSMTNRTDVAT
VNOIKALERYGADIVRSVPTMDAAEFKIKORVAVPIADIEPRILAKABYGV
DCRINPNGINBERIRIMVVDCAKDKITPIRIGNVNSLEKIDQEKIGEPPLAES
AMRHVDLDRLENFQKRVSVKASDVLAESYRLAQIDQPLHGLTTEGGASGAV
KSAIGLILSEBIDTLRVSLAADPVEIKVGDILKISIRSGINFIACPTCSRO
EFPIVIGTVALBQRLBDITPMDSIIGCVNNGEGALVSTLIGVGNKKSGLYEDGV
RKORLDNNMDIOLBARIRAKASQLDEARRIDVQVGKIT"
complement(7669..8682)
/gene="yfgA"
/locus_tag="SF2562"
complement(7669..8682)
/gene="yfgA"
/locus_tag="SF2562"
/notes="Residues 1 to 337 of 337 are 99 pct identical to
residues 1 to 337 of a 337 aa protein from Escherichia
coli O157:H7 ref: NP_311405.1"
/codon_start=1
/translation="putative membrane protein"
/product="putative membrane protein"
/protein_id="AAM4062.1"
/db_xref="GI:24052936"
/translation="MNTTEATHDONEALTGARRLNAREQGLSQOAVAERLCKVSTV
RDLEEDKAPADLASTFLRGYIRSYARLVALEBELLEKQAPLRAKAPVQGSFL
GKRKKRDKGMLMTLVLVFLVIGLSGAMMMDHKAQOEITTMADQSALENSNSO
GQSVPLNTSTTDDPATSTPQASVDITATNTOTPAVYAPAPAVDPOONAVVSSQANV
DTAATPVPATATTPDGAAPLPDQAGVTTTAAADBNALVMTFADCMLEVDATGCKLF
```

Alignment Scores:  
Pred. No.:  
Score:

1.2e-133 Length: 10225  
1855.00 Matches: 370

```
Percent Similarity: 99.73% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 1  
Query Match: 99.41% Indels: 0  
DB: 1 Gaps: 0  
US-09-921-992-50 (1-372) x AB015271 (1-10225)  
Qy 1 MetHisASnGLAAlaProIleGlnArgGlySerThrArgIleThrValGlyAsnVal 20  
Db 7642 ATGCATACCGAGCTCCATTCATCGTAAGAAATCAACGATTTACCTGGGAATGTG 7583  
Qy 21 ProIleGlyASpGlyValaProIlealaValaGlnSerMetThrAsnThrArgThrAsp 40  
Db 7582 CGGATGGGAGATGTCTCCATCCGCGGATCAATGATCAACCAATCGGATACAGAT 7523  
Qy 41 ValGluAlaThrValaAsnGlnIleLysAlaLeuGluArgValaGlyAlaAspIleValaArg 60  
Db 7522 GTCGAAGCAACGGTCATCAATCAATCAAGGCGCTGGAAAGCGGTTGGCCTCATATCGATCGGT 7463  
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValaAsn 80  
Db 7462 GTCTCCGTCACCGATGAGACGGCGGCAAGACGTTCAAACTCATCAACGCGGTTTAC 7403  
Qy 81 ValProLeuValaAlaAspIleHisPheAspTyrArgIlealaLeuLysValaAlaGluTyr 100  
Db 7402 GTCCCGCTGTGGCTGACATCCACTTCGATTCGATTCGCGTGAAGAGCGAATAC 7243  
Qy 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
Db 7342 GGCCTGATGTCCTGCTGATTTAACTCGTGAATATCGTAATGAAAGAGGTATTCGATG 7283  
Qy 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140  
Db 7282 GTGTTGACTGTGGCGGCAATAAACATTCGATTCGATTCGATTCGATTCGATTCGATTCG 7223  
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
Db 7222 CTGGAAAGATCTGCAAGAAAGATGAGCAACGACCGCGGATTCGATTCGATTCG 7163  
Qy 161 AlaMetArgHisValaAspHisLeuAspArgLeuAsnPheAspGlnPheLysValaSerVal 180  
Db 7162 GCATCGCATGATGATCATCATCATCGATCCGTCGATCGATCGATCGATCGATCGATCG 7103  
Qy 181 LysAlaSerAspValaPheLeuAlaValaGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
Db 7102 AAGGCTCTGACGCTCTTCCTCGCTGTTAGCTTATTCGTTGCTGCAAAACGATCGAT 7043  
Qy 201 GlnProLeuHisLeuGlyIleThrArgIleAlaGlyValaArgSerGlyValaLysSer 220  
Db 7042 CAGCGCTGATCTGGGATCAACGAAAGCGGTGTCAGACGACGCGGCGAGTAAATCC 6983  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyAspThrLeuArgValaSerLeu 240  
Db 6982 GCCATGCTGTTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6923  
Qy 241 AlaAlaAspProValaGluGluIleLysValaGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 6922 GCGGCGGATCGCTGCAAGATCAAGATCGGTTTGAATTTTGAATTCGCTGCTGATC 6863  
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
Db 6862 CGTTCCGAGGAGATTAATCATTCATCGCTCGCGGACCTGTTCCGCTGAGGAATTTGATGTT 6803  
Qy 281 IleGlyThrValaAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
Db 6802 ATCGGATCGGATTAACCGCTGCGCAACCGCTGGAAGATATCATCTCGAAGAGCGATT 6743  
Qy 301 SerIleIleGlyCysValaValaAsnGlyProGlyGluAlaLeuValaSerThrLeuGlyVala 320  
Db 6742 TCGATTAATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6683  
Qy 321 ThrGlyIleAsnLysLysSerGlyLeuTyrGluAspGlyValaArgLysAspArgLeuAsp 340
```

Qy 321 ThrGlyAlaAsnYsLysSerGlyLeuTyGlnAspGlyValAlaArgLysAspArgLeuAsp 340  
Db 187450 ACTGGCGCAACAAGAGAGCGGCTCTATGAGATGCGCGCCCAAGACCGCTGTGAC 1873391  
Qy 341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgLysAlaValAspGlnLeuAsp 360  
Db 187390 AACCAACATATGATCGATCGATCGAGGCGGCACTTGTGCGAAGCCAGCTGAGC 1873331  
Qy 361 GlnAlaArgArgIleAspValGlnGlnValGlnLys 372  
Db 187330 GAAGCGCTGCAATTGACGTTGACGACGTTGAAAA 187295  
RESULT 11  
AE015271/c  
LOCUS  
DEFINITION Shigella flexneri 2a str. 301 DNA linear BCT 18-OCT-2002  
ACCESSION AE015271 AE005674  
VERSION AE015271.1 GI:24052928  
KEYWORDS  
SOURCE  
ORGANISM Shigella flexneri 2a str. 301  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Shigella.  
REFERENCE  
AUTHORS 1 (bases 1 to 10225)  
Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H.,  
Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X. B., Zhang, J. Y., Yang, G. W.,  
Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, A. L., Gao, Y. S., Zhu, J. P.,  
Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L.,  
Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.  
Genome sequence of Shigella flexneri 2a: insights into  
pathogenicity through comparison with genomes of Escherichia coli  
K12 and O157  
Nucleic Acids Res. 30 (20), 4432-4441 (2002)  
JOURNAL  
PUBMED 2 (bases 1 to 10225)  
12384590  
AUTHORS Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B.,  
Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y.,  
Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y.,  
Lu, W. C., Qiang, B. Q., Wen, Y. M. and Hou, Y. D.  
Direct Submission  
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry  
of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P. R.  
China  
FEATURES  
source Location/Qualifiers  
1..10225  
/organism="Shigella flexneri 2a str. 301"  
/mol\_type="genomic DNA"  
/strain="301"  
/serotype="2a"  
/db\_xref="taxon:198214"  
148..1440  
/gene="xseA"  
/locus\_tag="SF2555"  
148..1440  
/gene="xseA"  
/locus\_tag="SF2555"  
/note="Residues 1 to 430 of 430 are 97 pct identical to  
residues 27 to 456 of a 456 aa protein from Escherichia  
coli K12 ref: NP\_417004.1"  
/codon\_start=1  
/transl\_table=11  
/product="exonuclease VII, large subunit"  
/protein\_id="AA044055.1"  
/db\_xref="GI:24052929"  
/translation="MGQWISGIEISNFTOPASGHYFTLKDDTAOVRCAFRNSNRV  
TTPHGGQVLYRANTLYRPGDYQIVESMPAGEGLQKYEQLKAKLOAECLD  
QQKRPSPAHGCVITSTKTAALDILHYLKRPPSLPVITPTYSVQDDAPGQIV  
RALEIANORVECVLIVRGSGSLEDLMSFNDERVAIPASRIPIVSAEGHEDVTI  
ADVAIDLRAITPSAAAEVVSRRNQELRQVSRQLLEAMDYLLNRRRRPQIHR  
LQOQHPLRLAQOTMLERLQKMSFALENQLKACQCOQRLTOLNQQNPQPKIART  
QTRIQOLEYRLAIELRAQLSATRERFGNAVTHLEAVSPILTAAGYSVTATADGVLYK

gene complement(1437..1688)  
/gene="yfgJ"  
/locus\_tag="SF2556"  
complement(1437..1688)  
/gene="yfgJ"  
/locus\_tag="SF2556"  
/note="Residues 1 to 83 of 83 are 97 pct identical to  
residues 1 to 83 of a 83 aa protein from Escherichia coli  
K12 ref: NP\_417005.1"  
/codon\_start=1  
/transl\_table=11  
/product="orf, conserved hypothetical protein"  
/protein\_id="AA044056.1"  
/db\_xref="GI:24052930"  
/translation="MNVGEMATGIGIMELHCPKCOHVLDDQNGHARCPGCEFIEMKA  
LPPDCHQPLQVLKACGAVYFCOHGILSKRVERVLA"  
complement(1721..3232)  
/gene="yfgK"  
/locus\_tag="SF2557"  
complement(1721..3232)  
/gene="yfgK"  
/locus\_tag="SF2557"  
/note="Residues 1 to 503 of 503 are 99 pct identical to  
residues 1 to 503 of a 503 aa protein from Escherichia  
coli O157:H7 EDL933 ref: NP\_289064.1"  
/codon\_start=1  
/transl\_table=11  
/product="putative GTP-binding factor"  
/protein\_id="AA044057.1"  
/db\_xref="GI:24052931"  
/translation="MRCLMTYKNEALNMVPAVALVGRPNVSKSTLFRRLTRDALVA  
DPPGLRDRKDAEIEGREFICIDTGGIDGNEDEQVMAQSLLAIEADVLFMV  
DARGLMPADPEAIAXHLSREKPTCTIDGGIDGNEDEQVMAQSLLAIEADVLFMV  
GRGVSILHEVILPMWEDLAPQEVDEDEMYAQFPAEENGESEEDDPQSLPIKL  
ALVGRPNVSKSTLFRRLTRDIEERVYVDMQGTTRDSIYIMERGRVYLIDTAGVGR  
KITDVAKEKSVLKTQLAIEDANVVMVIDAREGISDQSLGLFINSGRSLVIVN  
KMDGSOEVEKQVETLDFRGLFIDPARVHTISALHSGGVNLFESVRAVDSSTRV  
GTSMLTRITWMAVEDHDPVLRGRVYKLYAHAGGVNPIVTHGNOVDLPDSYKRY  
LMNYPRKSLDWNQSPRIQPKGENDPANKRNTLTPTQNRKKRLMKIMKK"  
complement(3311..4489)  
/gene="yfgL"  
/locus\_tag="SF2558"  
complement(3311..4489)  
/gene="yfgL"  
/locus\_tag="SF2558"  
/note="Residues 1 to 392 of 392 are 99 pct identical to  
residues 1 to 392 of a 392 aa protein from Escherichia  
coli K12 ref: NP\_417007.1"  
/codon\_start=1  
/transl\_table=11  
/product="putative dehydrogenase"  
/protein\_id="AA044058.1"  
/db\_xref="GI:24052932"  
/translation="MQLRKLLPLGLSVTLTLLSGSLFNSBEDVVKMSPLPTVENQFPT  
TAMSTSVSGSIGNFYSNKLPLADLVVAAARAGLVKLNDDGKEIYSLSAEDG  
WRSKEPALISGCVTVSGHYVYTSKRAQVYALANTAGTAAKQTKVAGALSAPVSDG  
LYIHITSNQLQALNBADGAVKRTVNLDPSPSLTSGSAPATAFGAAYVGGNGRISA  
VIMEQGMWQORISOATGSTEIDRLSDVDTTPVVANGVFLAYNGNLTALDRLSGQ  
IMWKEBELSVNDPFIVDGNRIYLVDDNRVMAITIDGGVTLMAOSDLHLRLTSPVLYN  
GMLVVDSESGVLYHVINVEDGRFVAQOKVDSGFQTEPVAADKLLIQARDGVYSITR  
"  
complement(4500..5120)  
/gene="yfgM"  
/locus\_tag="SF2559"  
complement(4500..5120)  
/gene="yfgM"  
/locus\_tag="SF2559"  
/note="Residues 1 to 206 of 206 are 98 pct identical to  
residues 1 to 206 of a 206 aa protein from Escherichia  
coli K12 ref: NP\_417008.1"  
/codon\_start=1  
/transl\_table=11

```
/protein_id="AA081323.1"
/db_xref="GI:26109121"
/translation="METPNSLFWVSPLLGLVLFPMVLAAGFIDSIAGGGGLTTPALM
AAGMSPANALATKTLQACGSGISATITFIRKRVYSIDQELNTAMTFVSGMSALLVO
YVADVLRQTLPLVLTCTGLFLMPLGLEDQRKRYGLFPLIAGGCVGFIDGFEFG
PAAGSFALAFTVLCTGFENLAKATAHAKLATNATNIGLLFLIIGKVTATGFMVLVG
QFGLAMGSRILVLSKQGLIRPMIIVISAVMSAKLIYDSHGEILHMLGN"
complement(5338..6162)
/gene="mepa"
/locus_tag="c2874"
/complement(5338..6162)
/gene="mepa"
/locus_tag="c2874"
/BC_number="3.4.99.-"
/function="enzyme; Murein sacculus, peptidoglycan"
/note="Escherichia coli K-12 ortholog: b2328; Escherichia
coli O157:H7 ortholog: z3591"
/codon_start=1
/transl_table=11
/product="Penicillin-insensitive murein endopeptidase
precursor"
/protein_id="AA081324.1"
/db_xref="GI:26109122"
/translation="MNKTAIALALALASASLATPWOKITOPVPSAQSISGSPNGC
IVGADTLPIQSEHYQVWRTDQRYFGHPDLWFIQRLRSVSNLGMGTVLIGDMGPA
GSRFNGSHASHQGLDVDFIOLPKTRMSTAOILRQALVSRDGHVPAIMKPEI
FSIILKAAQKDVTRIIVNPAIKQOCLDAGTDRDMLRKVRPMRQRAHNVLRCPA
DSLCEDDQLPPEDGCGAELQSWFEPKPTTKPEKKTPLPSPQALLDEHVI"
complement(6166..7254)
/gene="aroc"
/locus_tag="c2875"
/complement(6166..7254)
/gene="aroc"
/locus_tag="c2875"
/BC_number="4.2.3.5"
/function="enzyme; Amino acid biosynthesis: Chorismate"
/note="Escherichia coli K-12 ortholog: b2329; Escherichia
coli O157:H7 ortholog: z3592"
/codon_start=1
/transl_table=11
/product="Chorismate synthase"
/protein_id="AA081325.1"
/db_xref="GI:26109123"
/translation="MMAGNTIGOLFVTTFGESHGLAGCIVDGPVGPILTEADLOH
DURRRRGTSRYTTQRRRPPQVKTLSVFGVTTGTSIGLIENTDROSDYRAXDV
FRGHADYTYEQKIGLRDYGGRSSARETAMVAAGAIKATLAEKFGLEIKCLTQ
MGDIPLEIKDMSOYEONPFCPPDDKIDALDELRLAKSGDSIGAKVTAVASGVPAQ
LGEVFPRLADIAHMLMSINAVKVEIGDFVALRGSQNRDEITKDFQSNHAG
ILGSSGQOI IAHMLKPTSSITVPGRTINRFGEVEMITKGRHDPVGRVAVPIAE
AMAIVIMDLRQANADVKTDIPRM"
complement(7286..8551)
/gene="yfcB"
/locus_tag="c2876"
/complement(7286..8551)
/gene="yfcB"
/locus_tag="c2876"
/BC_number="2.1.1.72"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2330; Escherichia
coli O157:H7 ortholog: z3593"
/codon_start=1
/transl_table=11
/product="Hypothetical adenine-specific methylase yfcB"
/protein_id="AA081326.1"
/db_xref="GI:26109124"
/translation="MAGIGLFLNQLYMHFADEFTLRSVDRVLLNLASPRHOLAKOR
LVLLAKCCFLHLSFSISNPTVYPTSTIDRFKRIPPSPRVNADLSSSWCTSRN
YCHALREYVDKIFVDEAVNLOTIDMLRWSVSPSAAINWGHGNDPMDVAVOLV
LBSLYLIDIDEDMRTARLTSEKGRIVERIYIRVNRICRIAPVYLTNAAKCHGEFYD
BRVAVPSPIGELINNKFAGLISKQPHILDMCTGSGCILAICAYAPAEAEVDVDS
PDALVAEONTEERHGLHNVTPIRSDFRDLPKQYDLITVNPYVAEEMSDIPNYS
RHEBELGLASGTDLKLTIRILGNADYLDADVLTICEVNSVHMLMEQYDPVFTWL
```

```
Alignment Scores:
Pred. No.: 1,47e-132 Length: 300099
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-921-992-50 (1-372) x AE016764 (1-300099)
QY 1 MethiAsnGlnA1aProIleGlnA1aGlySerThrArgIleTyrValGlyAsnVal 20
Db 188410 ATGCATTAACGAGCTCCAAATTCACCTAGAAATCAACGATTTACCTGGAAATGTC 188351
QY 21 ProIleG1A5pG1A1aProIleA1aValG1aGlnSerMetThrAsnThrArgThrAsp 40
Db 188350 CCATTGGCGATGTCCTCCATCGCGTACATTCATGACAAATCCGCTACGACAGAC 188291
QY 41 ValGluA1aThrValAsnGlnIleLysA1aLeuGluArgValG1A1aAspIleValArg 60
Db 188290 GTGGAAGCAACGGTCAATCAATCAAGCGCTGGAAAGTGTGGCGCTGATTCGTCGT 188231
QY 61 ValSerValProThrMetAspA1aA1aGluA1aPheLysLeuIleLysGlnValAsn 80
Db 188230 GTCTCCGTACCGACGATGACCGCGCAGAAAGCTTCAAACTCATCAACGCGATTAC 188171
QY 81 ValProLeuValA1aAspIleHisPheAspTyrArgIleAlaLeuLysValA1aGluTyr 100
Db 188170 GTCCCTCGTGTGCTACATCCATTCGATACGATACGATTCGCTGAAAGTACGGAATAC 188111
QY 101 GlyValAspCysLeuArgIleAsnProG1A1aG1A1aAsnGluGluArgIleArgMet 120
Db 188110 GCGCTGATTCGTGCTGATTAACCTGGCAATATCGTATGAAGCGCTATTCGATG 188051
QY 121 ValValAspCysA1aArgAspLysAsnIleProIleArgIleG1A1aAsnA1aGlySer 140
Db 188050 GTGGTTGACGTGTCGCGCGATTAACCAATTCATTCGATTCGCTTAATGCGGATTCG 187991
QY 141 LeuGluLysAspLeuGlnGluTyrArgIleProThrProGlnA1aLeuLeuGluSer 160
Db 187990 TTGGAATAAGATTCGAAGAAAGATGAGGAAACCAAGCGCGGCGCTTCGGAATCT 187931
QY 161 AlMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 187930 GCATCGCATTCATTCATCATTCGATCGCTGAATTCGATTCGATTCGATTCGATTC 187871
QY 181 LysAlSerAspValPheLeuA1aValGluSerTyrArgLeuLeuA1aLysGlnIleAsp 200
Db 187870 AAAGCGCTGACGCTTCCTCGCTGTTGAGTTCATTCGCTTCGCAAAACCAATCGAT 187811
QY 201 GlnProLeuHisLeuGlyIleThrGluA1aGlyValA1aArgSerGlyValA1aLysSer 220
Db 187810 CACCCGCTGATTCGAGGATCACCGAAGCGGCTGGGCGGAGCGGCGGAGTAATTC 187751
QY 221 AlaIleG1A1aLeuG1A1aLeuLeuSerGluGlyIleG1A1aAspThrLeuArgValSerLeu 240
Db 187750 GCTATTCGTTTAAAGTCTGCTGCTGCTGATGAGCATCGGCAACGCTGGCGATCGCTG 187691
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 187690 GCGGCGATCGGATCGAAGAGATCAAAAGTGGTTTCATTTTGAATCGCTCGCATC 187631
QY 261 ArgSerArgGlyIleLeuPheIleA1aCysProThrCysSerArgGlnGluPheAspVal 280
Db 187630 CGTTGCGCGGATCAACTTCATTCGCTCGCAACCTGTCGCTGCGGATTCGATTCG 187571
QY 281 IleG1A1aThrValAsnA1aLeuGluGlnArgLeuGluLysAspIleIleThrProMetAspVal 300
Db 187570 ATCGATTCGTTTAAAGCGCTGAGCAACGCTGGAAGATTCATTCATTCGATTCGATTC 187511
QY 301 SerIleIleG1A1aCysValValA1aGlnGlyProGlyGluA1aLeuValSerThrLeuGlyVal 320
Db 187510 TCGATTATTCGCTGCGTGGAAATGCGCCAGGTGAGCGCTGATTCATTCATTCGCTG 187451
```

QY 301 SerIleIleGIcYsvaValaangIyProGIyGluIalaleuValserThleuGIyVal 320  
Db 279446 TCGATTATCGCGTGGCTGGTAATGCGCCGGGTGAGGCGCTGGTTCTACACTCGGGTC 279387  
QY 321 ThGIGIYAsnIyIyIySserGIyLeuTYrGIYAspGIYValIArgIyAspATgLeuAsp 340  
Db 279386 ACCGGGGCAACAGAAACCGGCTCTTATGAAGATGCGCTGGCAAGACCGTCTGGAC 279327  
QY 341 AsnAspMetIleAspGInleuGluIalAArgIleArgIalIyAsIasergInleuAsp 360  
Db 279326 AACACGATATGATTCACACCGCTGGAGCGACCGCATTCGTCGGAAGCCAGTCACTGGAC 279267  
QY 361 GluIalAArgArgIleAspValGInGInValGluIyS 372  
Db 279266 GAAGCCGCTGGAATTGACGTTGACGAGGTTGAAAAA 279231  
RESULT 10  
AE016764/c 300099 bp DNA linear BCT 24-DEC-2002  
LOCUS Escherichia coli CFT073 section 10 of 18 of the complete genome.  
DEFINITION AE016764 AE014075  
ACCESSION AE016764.1 GI:26109116  
VERSION  
KEYWORDS  
SOURCE Escherichia coli CFT073  
ORGANISM Escherichia coli CFT073  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1 (bases 1 to 300099)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.  
Extensive Mosaic Structure Revealed by the Complete Genome Sequence  
of Uropathogenic Escherichia coli  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)  
12471157  
2 (bases 1 to 300099)  
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,  
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,  
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,  
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.  
Direct Subassembly  
Submitted (20-JUN-2002) Genetics Laboratory, University of  
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA  
LOCATION/Qualifiers  
1. 300099  
/organism="Escherichia coli CFT073"  
/mol\_type="genomic DNA"  
/strain="CFT073"  
/db\_xref="taxon:199310"  
complement(163..1383)  
/gene="fabB"  
/locus\_tag="c2869"  
complement(163..1383)  
/gene="fabB"  
/locus\_tag="c2869"  
/FC\_number="2.3.1.41"  
/function="enzyme; Fatty acid and phosphatidic acid  
biosynthesis"  
/note="Escherichia coli K-12 ortholog: b2323; Escherichia  
coli O157:H7 ortholog: z3586"  
/codon\_start=1  
/transl\_table=11  
/product="3-oxoacyl-[acyl-carrier-protein] synthase I"  
/protein\_id="AA081319.1"  
/db\_xref="GI:26109117"  
/translation="MKRAVITGIGIVSIGNNQOEVLASIRGRSGITRPSOEIKDSGM  
BSHWGNVLTDTTGLIDRKAIVFMSDASTIYAFISMEOATADAGLSPEAYONPRCLT  
ASSGGSPPFOVFGADAMGRPGSLKAVGVYVTKMASVSAKLATPFKIHGYNYSIS  
SACATSAHCTIGNAVEQIQKODIVFAGGEBELCWMACEFDMAGLSTKYVDTPERKA  
SRTYDAHDFGVYIAGGGMVVELELHARGAHYAEIVGAGTSDGADAPSGEG  
ARCKMAMHAGVDTPIDYLSHGTSTPVDVKEALAIREVFGDKSPAIATATAMGHS  
LCAAGVQEAIVSLMLHEGFIAPISINIEELDQAGLINTVETTDRELTTWNSNGCF

gene  
1482..3548  
/gene="yfcK"  
/locus\_tag="c2870"  
1482..3548  
/gene="yfcK"  
/locus\_tag="c2870"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2324; Escherichia  
coli O157:H7 ortholog: z3587"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yfcK"  
/protein\_id="AA081320.1"  
/db\_xref="GI:26109118"  
/translation="MRSLTIAKIEPPRYRTYVYKHSIQPANIIEFNAEGFPVSDF  
DDVYFSNDGLEETRYVFLGSHLEKPEFHPRLFVAVESGFGTGLNLTLMQADQ  
FREARPOQLRIHFISEKFPRLRADLALAHQWMLDPMAEQLQAWPLPLPGGR  
LLIDEGHVTLDLMPGIDINELTQLDLSNQKVDAMPIDGFAKAPNDMMTOMLFNMA  
RLARGSTLATPTASGFVRGRIOEAGFTWQKRGFKRKBMLCGVVEQTLPPFCSTPM  
FNRMSNQEAAIIGGFIASALISALLRGRQVTLYCDAEPALGASQKRAQALVPL  
ISKHPBALNRPFSNAFTFARRFDLLPVKFDHDCGYTOLGMDKSQKHAQMLSHLD  
PAELVAEAVANVEQITGVATNCSGITYPQGLCPAELTRNVLBLAQOQGLQIHQ  
QLQDLSRKDDGMLNFPAGDPAHTSHSVVLANGHQSFGSQTSSLPYVSAGOVSHIPT  
TPELAKQVLCYDGYLTPQNPANQHHICGAYHSGSEETAYSDQOONROBLIDCF  
PQVONAKEDVDSGEARPGVCATDHLPMVGNVDPYDATIVEVASIPAKREAVSAP  
VYDDLPMFAALOSRLCSNPLCAETLAAQMSSEPIPMDSATLAALNPNLWVKLLKG  
KAVKXG"  
complement(3669..3947)  
/gene="yfcL"  
/locus\_tag="c2871"  
complement(3669..3947)  
/gene="yfcL"  
/locus\_tag="c2871"  
/note="Escherichia coli K-12 ortholog: b2325; Escherichia  
coli O157:H7 ortholog: z3588"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yfcL"  
/protein\_id="AA081321.1"  
/db\_xref="GI:26109119"  
/translation="MIAEFERSLIALIDGMVDHSDDELFPASGYRGLHTLAIALES  
GDHSAQAVHTVTSQLEKAIAGELSPRDQLVTDMMHLLFQASQ"  
complement(3981..4529)  
/gene="yfcM"  
/locus\_tag="c2872"  
complement(3981..4529)  
/gene="yfcM"  
/locus\_tag="c2872"  
/function="putative"  
/note="Escherichia coli K-12 ortholog: b2326; Escherichia  
coli O157:H7 ortholog: z3589"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yfcM"  
/protein\_id="AA081322.1"  
/db\_xref="GI:26109120"  
/translation="MNSTHYEQILIEINSCFADEFNTRLIGDDEPIYLPADAEPY  
NRIVPAHGFYSAIHIEISHWCJAGARDDLVDPGWYCPDGDQAQTOSQFEDVEVQ  
ALDMLFCVAGYPPFVSCDNLGDFEPDRVVFQRRVHAQVMDYLANGIPERPARFIKA  
LQNYHTPELTLEQFPWPEALN"  
complement(4529..5338)  
/gene="yfcA"  
/locus\_tag="c2873"  
complement(4529..5338)  
/gene="yfcA"  
/locus\_tag="c2873"  
/function="putative structure"  
/note="Escherichia coli K-12 ortholog: b2327; Escherichia  
coli O157:H7 ortholog: z3590"  
/codon\_start=1  
/transl\_table=11  
/product="Hypothetical protein yfcA"



```

gene
  4420..6048
  /gene="ECs3126"
  4420..6048
  /gene="ECs3126"
  /note="similar to GlpA ECOLI gi|1788574 percent identity
  100 in 542 aa (Conserved in E.coli K-12)"
  /codon_start=1
  /evidence=not_experimental
  /transl_table=1
  /product="anaerobic sn-glycerol-3-phosphate dehydrogenase
  large subunit"
  /protein_id="BAB36549.1"
  /db_xref="GI:13362596"
  /translation="MKTRDSQSSDVIIIGGAGAGIARPCARGLRVILVERHDAT
  GATRNHGLHSGARIATVDAESARECTISNQILKRIARCVPTNGPLTFLEDDIS
  FOATFIRACEAGISABAIIDPOQARIIEPVNPAI.GAVVPDGVADPFLTANMLD
  AKHGAVILTAHEVTGLIREGATVCGVRNHLTGEQALHAAVVAAGIMQCHIAE
  YADLRIMFPAKSLIMDRINOHVNRKPSDAILVPGDTISLIGTTSRIDYN
  EIDNRYTAEVDIILREGEKLA.PVMAKRIILRAYSQVRLVSDDDPSGRNRYGIV
  LIDHAREDDGDRITITGKLTMRPLMAEMWATDAPVCRKLTGPTTADALGSGDEP
  AETTLKRVISLPAPLRGSANYRHGDRTPAWLSGRHLRLVCECAVTAEEVAYVEN
  LVNLSLIDLRRTTRVWGTCGELACRAAGLLQRFVNTTSAQSIQSLTFINRWKG
  VQPIAWGDALRESEFTRWVYQGLCGLEKEKQDAL"
  6038..7297
  /gene="ECs3127"
  6038..7297
  /gene="ECs3127"
  /note="similar to GlpB ECOLI gi|1788575 percent identity
  98 in 419 aa (Conserved in E.coli K-12)"
  /codon_start=1
  /evidence=not_experimental
  /transl_table=1
  /product="anaerobic sn-glycerol-3-phosphate dehydrogenase
  membrane anchor subunit"
  /protein_id="BAB36550.1"
  /db_xref="GI:13362597"
  /translation="MRPDTIMGGGLAGLIGLOLKHGRCAIVTRGOSALHPSGGS
  LDLSHPDQAVTDIHSGLESLRQAPARPYSLGQRVLDAQAOALIASGQOL
  QSSVELPHQRIPLGLRSTWLSPEVPVPLPAKICVVGISGLMDFOHLLAAASIR
  QSLVAETAIELEPELDVLRNNATEFNAVIAFLDNEENWPLLDALFVANCTEMI
  LMPACFLADDKLMRWLNELPCSLMLPTLPPSVLGIQLONOLQRFVQCGVMMG
  DEKAVTCKGVNIEWIRNHADILPRPRAVLAGSFFSGGLVAEENGIREFIIGLD
  VLDTATRGWYKGFAPAPQWQOQGYTDTLAPRSQAGCTIENLFTIGSVLGFPDPIA
  QCGGCGVCAVSAHHAQQLAKRAGGQO"
  7294..8484
  /gene="ECs3128"
  7294..8484
  /gene="ECs3128"
  /note="similar to GlpC ECOLI gi|1788576 percent identity
  100 in 396 aa (Conserved in E.coli K-12)"
  /codon_start=1
  /evidence=not_experimental
  /transl_table=1
  /product="anaerobic sn-glycerol-3-phosphate dehydrogenase
  K small subunit"
  /protein_id="BAB36551.1"
  /db_xref="GI:13362598"
  /translation="MNDTSFENCICKTCTTACPVSRVNPFGPKQAGPGRRLK
  DGAFLDALKYCINCKCEVAECPSDYKIGIQRARAKYTPRPSLNFVLSHDDMG
  SVSPFAPIVNTATSLKPVROLDAALIKHRLPLPSGTRRWRYVAQAQOYK
  DOVAFFHGFVNHNHQLGKDLIKVLANMGTGLSKESKCCGPIILANPDMARQ
  AIRNVEIRRAVGVKGIPIVATISSTCTFALRDEYPRVIANDKGLRDLATRMWR
  KLDBGKTLPLKPLKAVHATPCHEKMGWTLTLELRNIPGLLETLVDSQCGGLAG
  TYGFKENYPTSAIGAPLFRQIESGADLVVDCTCKQIEMSTSLRCEHPITLLA
  QALA"
  8477..9603
  /gene="ECs3129"
  8477..9603
  /gene="ECs3129"
  /note="similar to yPAD_ECOLI gi|1788577 percent identity
  94 in 303 aa but differs at C-ter (Conserved in E.coli

```

```

Alignment Scores:
Pred. No.: 1,45e-132 Length: 296827
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-921-992-50 (1-372) x AP002561 (1-296827)
QY 1 MethiAsnGlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnVal 20
  280346 ATGCATACACGAGCTCAATTCAAGCTAGAAATCAACGATTTTACGTTGGGAATGTC 280287
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
  280286 CCGATTGGCGATGCTCTCCATCGCGCATGACATCAATCAAGTCCATCAATCGCGTACGACAGAT 280227
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
  280226 GTCGAAGCAACGCTCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATATCGCCG 280167
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
  280166 GTCTCCGTCACCAAGATGACGCGGCAAGAGCTTCAAACTCATCAACAGCAGTTAAC 280107
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
  280106 GTCCCGCTGATGCTGACATCCACTTCGATTCGATTCGCGTGAAGTACGGAATAC 280047
QY 101 GlyValAspCysLeuArgIleLeuAsnProGlyAsnIleGlyAsnGluGluArgMet 120
  280046 GGGGTGATGCTGCTGCTGTTTAACTCGCAATGGAATGAAAGCGGTATTCGATG 279987
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
  279986 GTGCTTGACTGTCGCGGCATTAATAAATTCGATCCGATTCGTTACGCCGATCG 279927
QY 141 LeuGluLysAspLeuGlnGluTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
  279926 CTGGAATAAATGCTGCAAGAAAGTATGGCAACGACCGCGGCGTTCGATCC 279867
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnAspGlnPheLysValSerVal 180
  279866 GCCATCGCGCATGTTGATCATCTCGATCCGCTGAACCTTGATCAAGTCAACGACGTCG 279807
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
  279806 AAGGCTCTGACGCTCTCTCGCTGTTGAGCTTATCGTTTGTGCGCAAAACAAATCGAT 279747
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValArgSerGlyValAlaLysSer 220
  279746 CAGCGCTGATCTGGGATCAACCGAGGCGGCGCGCGACGCGGAGAGTAAATCC 279687
QY 221 AlaIleGlyLysGlyLeuLeuLeuSerGlyIleGlyAspThrLeuArgValSerLeu 240
  279686 GCCATTGCTTGGTCTGCTGCTGCTGCTGCAAGGATCGGAGACGCTGCGCGATACCG 279627
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
  279626 GCGGCGGATCCGCTGCAAGATCAAGATCGGTTTGATTTTGAATTCGCTTCGATTC 279567
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280
  279566 CGTTGCGAGGATCAACTTCATCGCTGCGACCTGTTCCGCTGAGGAATTTGATGTT 279507
QY 281 IleGlyThrValAsnAlaLeuGluGluArgLeuGluAspIleIleThrProMetAspVal 300
  279506 ATCGGTACGATTAAACCGCTGAGCAACCGCTGGAAGATATCATCATCGATGACGTT 279447

```

Db	364	ATCGGTACGGTTAAAGCCCGTGGACGACCCCTGGAAAGATATATCATCTCCATGGACGTT	305
Qy	301	SerTIIeEGIcYsVaIvaIaSnGlyProGIgIuaIaLeuValSerThrIenGlyVal	320
Db	304	TGCAATTATCGCGCTCGGTGTGAATGAGCCCAAGGTGAGCGCTGGTTTCATCACTCGGCGTC	245
Qy	321	ThrgIyGIyaEnIyLysSerGIyLeuTyrGIuaaPgiYalArGIyAaSPaRgLeuSP	340
Db	244	ACGGCGGCGCAACAAAAAAAGCGGCTCTATGAAGATGGCGGCGGAAAAAGACCGTGTGGAC	185
Qy	341	AsnAaSPMeTIIeASpGIInIeugGIuaIaArgIleARgAlaYsAlASerGIInIeuaSP	360
Db	184	AACAACGATATGATGACACCACTGGAGACGCGCATTTCTGGCAAGACCACTGCTGGAC	125
Qy	361	GIuaIaARgArGIleASpYalGIInIyAlGIuIyS	372
Db	124	GAAGCGCGTCGAAATTGACGTTGACAGGTTGAAAA	89
RESULT 9	AP002561/c		
LOCUS	AP002561	296827 bp	DNA linear BCT 24-APR-2003
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 12/20.		
ACCESSION	AP002561	BA000007	
VERSION	AP002561.1	GI:13362592	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			

REFERENCE	5 (bases 1 to 296827)
AUTHORS	Omishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and Hayashi,T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail:kwen@gen-info.osaka-u.ac.jp), URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365, Fax:81-6-6879-2047)
COMMENT	genome project.
FEATURES	location/Qualifiers
source	1..296827
	/organism="Escherichia coli O157:H7"
	/mol_type="genomic DNA"
	/strain="O157:H7"
	/sub_strain="RIMD 0509952"
	/db_xref="taxon:83334"
	414..1484
	/gene="ECs3123"
gene	414..1484
CDS	/gene="ECs3123"
	/note="unknown"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="BAB36546.1"
	/db_xref="GI:13362593"
	/translation="MPVTSIHLSCNLIIONNKKLOPTELYKESOOTAKLEPFCDERLIK
	SSGRVYKSTSDSIANLAFPTKASDAISIKOKITVSYGFPDPIHGFVHVTYKK
	EPADYDEPTLTPTRGNNGLLSVSEKGFISORTINLHFEAININSHVKEN
	HNIVQSTPEYKISKINSDDPKYTKSSKESISQYSDQRIQTCTYILKNTWYKYN
	SSSERDYKFNSSKEIKNTYKQIVQAQKLNOMKYINKLSEGLIDIDTKLAPLINDSK
	FSTRDPFFAYSKOEKNIPTFDNSKFVDPKKEGIMVQHSNGLVITGKYCEGVQTA
	TCQOYDRLRYINIFPFPEKLE"
	complement (1708..2784)
gene	/gene="ECs3124"
CDS	complement (1708..2784)
	/gene="ECs3124"
	/note="similar to GMPQ_ECOLI gi 1788572 percent identity
	99 in 358 aa (Conserved in E.coli K-12)"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="glycerophosphodiester phosphodiesterase"
	/protein_id="BAB36547.1"
	/db_xref="GI:13362594"
	/translation="MKUTIKLSMALIMSTITWGGSAMADSNKITYIAHNGAGSYLE
	EHLLPAAVAVAQADYLEDLDLVTWKDHLVLHDHLDLVDTVDADRPDARQGR
	YALDFITDEIKSLKFTGEPDIENGKQVTPGRPMKSDPRVHTEBELLEFVGLN
	STKNGLIPEIKAPWPHOEGDIKATLELVKTYGTGDDKLVYLQCFPADELRKR
	KNLELRPMGMDNLVOLIAVTDNMTQOKOPDGSWVNVYDMPKPGMKOVAEADAO
	IGDYHMLLEBTSQGNKIKLTGMVQDAQOKLVVHPYTVASDKLPYTTDVNQLYDA
	YNKAGVNGLEFTDPDPAKVFLENKE"
	complement (2789..4147)
gene	/gene="ECs3125"
CDS	complement (2789..4147)
	/gene="ECs3125"
	/note="similar to GLPT_ECOLI gi 1788573 percent identity
	99 in 452 aa (Conserved in E.coli K-12)"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="sn-glycerol-3-phosphate permease"
	/protein_id="BAB36548.1"
	/db_xref="GI:13362595"
	/translation="MLSTFKPAPSHKARLPAAEIDPTYRLRMQIFIGIFGYAAYIVL
	RKNFALAMPYLVEQSGRDLGRLSGISLAVGFSKDIMSVSDRNSPRLVPLAGLIL
	AAAVMLFMGVWPATSSIAVFWLLFLCGWFOGMPPCGRTVHMSKQERGVISGL
	WNCANHVGGGIPLLFLIGWAMFNMDNALYMAFCAILVALFAPAMMRDPGSGGLE
	PIEYKNDPDDDYNEKABQELTKAQIFMQVVLNPKLNTYIAIANVFPYLLRYGLDMS
	PIYLVKVNPRALDSSMAIFFYVAGIPGLIGMMSDKVFRGRRGATGVFPMVLVTII
	ATIVKPKHNPPTVDMICMIVIGFLYGVMLIGHALBLAPKKAAGTAAAGFLGFG

CDS .. 3683)  
     complement(2529..3683)  
     /feature="yfgB"  
     /function="orf; Unknown"  
     /note="f384; 100 pct identical to yfGB\_ECOLI SW: P36979"  
     /codon\_start=1  
     /transl\_table=11  
     /product="orf, hypothetical protein"  
     /protein\_id="AAC75570.1"  
     /db\_xref="GI:1788865"  
     /translation="MSEQLVTPENVTYTKDGKINILDLNRQOREFFKDLGEKPRADQ  
     VMKMHVYCDDEWTDINKVLRGKLEKVEAIRPEVEORSDGFIKWLAVGQO  
     RRETVYIPDDADTICVSSQVCALECKCPQTAQGFRNIRVEIIGVWRKATYQ  
     AAKVTGQRITVVMNGEPLNLNNVPAMEIMLDPFGLSKRYTISTGCVRA  
     LRKLDGIDVALAISTLHAENDEIRDEIVPINKKNIETFLAVRRYLEKSNAGQVAT  
     LEYVMIDHNDETEHAHQLEALKDTPCKNLIIPNPPGAPYGRSSNSRIDRPSKL  
     MSYGFETIYIKRTGSGDDIDAACQLAGVDIDRTKTRLKRMQGEAIDIKAV"  
     complement(3778..3806)  
     /note="factor Sigma70; predicted +1 start at 2642391"  
     complement(3786..3814)  
     /note="factor Sigma70; predicted +1 start at 2642399"  
     complement(3833..4264)  
     /gene="ndk"  
     /note="synonym: b2518"  
     complement(3833..4264)  
     /gene="ndk"  
     /BC\_number="2.7.4.6"  
     /function="enzyme; Purine ribonucleotide biosynthesis"  
     /note="f143; 100 pct identical to NDk\_ECOLI SW: P24233"  
     /codon\_start=1  
     /transl\_table=11  
     /product="nucleoside diphosphate kinase"  
     /protein\_id="AAC75571.1"  
     /db\_xref="GI:1788866"  
     /translation="MAIERPFSIIRKNVAVNRVIGNTPAREAGFKIVGKMLHLIV  
     EAGRGVHADKPPDGIIVPMTSGPIVSVLEGENAVQHRRLCATNPANLACT  
     LRADYADSLTENGTCSDSVESAAREIYFVEGEGVCRR"  
     complement(4322..4350)  
     /note="factor Sigma70; predicted +1 start at 2642935"  
     complement(4413..6725)  
     /gene="pbpC"  
     /note="synonym: b2519"  
     complement(4413..6725)  
     /gene="pbpC"  
     /function="putative enzyme; Not classified"  
     /note="f770; This 770 aa ORF is 31 pct identical (43 gaps)  
     to 480 residues of an approx. 720 aa protein PBPf\_BAC5U  
     SW: P38050"  
     /codon\_start=1  
     /transl\_table=11  
     /product="putative peptidoglycan enzyme"  
     /protein\_id="AAC75572.1"  
     /db\_xref="GI:1788867"  
     /translation="WPRILTRGCWITLAAAPFLPLFLAAGADKLWPLPHEVNPARY  
     VVAQGTDLPRWADADGIMRYPVPTIEDVSPRYLALINYEERWFKIPGNPVSVA  
     AMQDTSGVSGSGSTLTMOYARLDPHPKTFGKIQLMALDLEHLSKRELTILY  
     LNRAFFGGLTIGIGASNAWYLCSPANISYEAMLAVALPQAPSLRPDSRPEAEAA  
     RKVLERMAVQVMSRDEOVKESREEPITLAROMPOLAIPSRWMLGKSSDKITTLT  
     DGLQRLREELAQWKKRGLPRSSSLAMIVDHTMRKRGVTSVDLNDSDRFCHVDMV  
     NSIRSGSVLKFVFTGVALDBGLHPASLDVDPRRTEDIRPNSFDSGFGFISWSA  
     LVRSNLPVQVLEAVYGRKPRFAKLRNNGGLPLPNGAAPVLSLIGGAKKLEDMVA  
     ATTAARHAKQAKLQDPDPLLEKSSGSAAMIIRIMADEQNPIDBALPRVVAL  
     AKRTSGYGRDAMAVGNARVIGITVRGDDGTPVQGFPAVPELLANNONILLS  
     RSANIPEDRPNSRVRGVIWVGQSGSITLPEGRGDRRLATITMLDGSOPITLLPEOG  
     INGIRFPTIMLDNGKRVAADEQARQENINWPLPIEPMILPASRRARVRLPPASTGCP  
     PIGHQAQVLTQVTRDGAITRLPGAELATLPQSSGAGGERWVFLNGEPLTERGRN  
     VTLHLTDKDDYQLTLMVDVQGLIATVRFVQ"  
     complement(6726..11687)  
     /gene="b2520"  
     complement(6726..11687)  
     /gene="b2520"  
     /function="orf; Unknown"  
     /note="f1653; f1653"  
     /codon\_start=1

Alignment Scores:			
Pred. No.:	2,366-134	Length:	13176
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	1	Gaps:	0
US-09-921-992-50 (1-372) x AB000338 (1-13176)			
QY	1	MethsAnGIuAlaProIleGlnArgArgysSerThrArgIleTyrValGIyAsnVal	20
DB	1204	ARGCAATCCAGGCTCCATTCAACGTAAGAAATCAACGTAATTACGTGGAAATGG	1145
QY	21	ProIleGIyAspGIyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp	40
DB	1144	CCGATGGCGAATGGTGTCCCATCGCGCTGACGTCAACCAATACCGTACGACAC	1085
QY	41	ValGIuAlaThrValLaenGIleLysAlaLeuGIuArgValGIyAlaAspIleValArg	60
DB	1084	GTCGAGCAACGGCTCAATCAATCAAGCGCTGAAACGGTTGGCGCTGATATGTCGCT	1025
QY	61	ValSerValProThrMetAspAlaAlaGlnAlaPheLysLeuIleLysGlnGlnValAsn	80
DB	1024	GTAATCCGATCCGACGATGAGACGGCGCAACCGTTCAACATCAACACAGAGTTAAC	965
QY	81	ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGIuTyr	100
DB	964	GTGCGCGTGGGGGTGACATTCATTCGACTATGCGATGGCGTGAAGATGCGGATAC	905
QY	101	GIyValAspCySLeuArgIleAsnProGIyAsnIleGIyAnGIuArgIleArgMet	120
DB	904	GGCGTCGATGTGTGCGGATTAACCTTGCGCAATATCGTATGAAGAAGCGTATTCGATG	845
QY	121	ValValAspCySAlaArgAspLysAsnIleProIleArgIleGIyValAsnAlaGIySer	140
DB	844	GTGGTGTGCTGGTGGCGGCAATMAAACATTCGATTCGGTATGGCGTTAACCGCGATCG	785
QY	141	LeuGIuLysAspLeuGIuLysTyrGIyGIuProThrProGIuAlaLeuLeuGIuSer	160
DB	784	CTGAAAAAGATCTCGCAAGAAATATAGGCAACCGACCGCGAGGCTTGCTGAATCT	725
QY	161	AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal	180
DB	724	GCCATGGGTATATGTGATCATCTCGATCGCTGAATCTTGATCACTTCAAAAGTCAGGGTG	665
QY	181	LysAlaSerAspValPheLeuAlaValGIuSerTyrArgLeuLeuAlaLysGlnIleAsp	200
DB	664	AAACGCTGACGCTCTTCCTCGCTGAGCTTATGTTGCTGGCAAAACAGATCGAT	605
QY	201	GlnProLeuHisLeuGIyLileThrGlnAlaGIyGIyValaLysSerGIyAlaValLysSer	220
DB	604	CAGCGGTGACATCTGGAGATCACCGAAGCGGTGTGTGGCGACGGGGGCGATMAATCC	545
QY	221	AlaIleGIyLeuGIyLeuLeuLysSerGIuGIyIleGIyAspThrLeuArgValSerLeu	240
DB	544	GCCATTTGGTTAAGTCTGCTGCTCTCTCGAAGGCAATCGGCACACGCTGCGGATTCCTG	485
QY	241	AlaAlaAspProValGIuGIuLileLysValGIyPheAspIleLeuLysSerLeuArgIle	260
DB	484	GCGCGCATATCGGTCGAAGAGATCAAAAGCGGTTCATATTTTGAATATCGCTGCTATAC	425
QY	261	ArgSerArgGIyLileAsnPhelIleAlaCyProThrCysSerArgGlnIlePheAspVal	280
DB	424	CGTTCGAGAGGATCAACTTCATGCTGCTGCCGACCTGTTGCGCTCAAGAAATTTGAATGT	365
QY	281	IleIyIleThrValAsnAlaLeuGIuGlnArgLeuGIuAspIleIleThrProMetAspVal	300

Db 424 CGTTCGGAGATACCACTTCATCGCTCTCCGACTCTTCGCTGACGAGATTGATGTT 365

Qy 281 ILGGLYThrValAsnAlaLeuGlnArgLeuGlnAspIleIleThrPrometAspVal 300

Db 364 ATGGATACATTAACCGGCTGGAGCAAGCGCTGAAATATCATCTCCGATGGAGCGTT 305

Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320

Db 304 TCATATTATCGGCTCGCGTGGTGAATGGCCCGGGTGGAGCGCTGGTTTCTACACTCGGCGTC 245

Qy 321 ThGVLGlyValAsnIleValSerGlyLeuTyrGluAspGlyValArgIleAspArgLeuAsp 340

Db 244 ACCGGCGGCAACAGAAACCGGCTCTTAAGAAGATGCGCTGGCAAAAGCCGCTGGAC 185

Qy 341 AsnAspAspMetIleAspGlnLeuGluValArgIleArgIleValValAspSerGlnLeuAsp 360

Db 184 AACACGATATGATGACCGAGCTGGAGGACCGCATTTGCTGCAAAAGCACTGACTGGAC 125

Qy 361 GUAAlaArgArgIleAspValGlnGlnValGluLys 372

Db 124 GAGCGCGTCAATTGACGTTTACAGACGTTGAAAAA 89

RESULT 8  
AE000338/ 13176 bp DNA linear BCT 01-DEC-2000  
LOCUS Escherichia coli K12 MG1655 section 228 of 400 of the complete  
DEFINITION Genome.  
ACCESSION AE000338 U00096  
VERSION AE000338.1 GI:1788862  
KEYWORDS  
SOURCE Escherichia coli K12  
ORGANISM Escherichia coli K12  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1 (bases 1 to 13176)  
Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,  
Blythe,R., Collins-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,  
Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,  
Mau,B. and Shao,Y.  
The complete genome sequence of Escherichia coli K-12  
Science 277 (5331), 1453-1474 (1997)  
9278503  
2 (bases 1 to 13176)  
Blattner,F.R.  
Direct Submission  
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459  
3 (bases 1 to 13176)  
Blattner,F.R.  
Direct Submission  
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,  
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.  
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:  
608-263-7459  
4 (bases 1 to 13176)  
Plunkett,G. III.  
Direct Submission  
Submitted (13-OCT-1998) Laboratory of Genetics, University of  
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA  
This sequence was determined by the E. coli Genome Project at the  
University of Wisconsin-Madison (Frederick R. Blattner, director).  
Supported by NIH grants HG00301 and HG01428 (from the Human Genome  
Project and NCHGR). The entire sequence was independently  
determined from E. coli K12 strain MG1655. Predicted open reading  
frames were determined using Genemark software, kindly supplied by  
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.  
30332 [e-mail: mark@ambler.gatech.edu]. Open reading frames that  
have been correlated with genetic loci are being annotated with CG  
Site Nos., unique ID nos. for the genes in the E. coli Genetic  
Stock Center (CGSC) database at Yale University, kindly supplied by

Mary Berlyn. A public version of the database is accessible  
(<http://cgsc.biology.yale.edu>). Annotation of the genome is an  
ongoing task whose goal is to make the genome sequence more useful  
by correlating it with other data. Comments to the authors are  
appreciated. Updated information will be available at the E. coli  
Genome Project's World Wide Web site  
(<http://www.genetics.wisc.edu>). \*\*\* The E. coli K12 sequence and  
its annotations are periodically updated; this is version M54. No  
sequence changes. Annotation updates: updated gene identifications  
and products; all new functional assignments courtesy of Monica  
Riley; added promoters, protein binding sites, and repeated  
sequences described in reference 1. The unique numeric identifiers  
beginning with a lowercase 'b' assigned to each gene (protein- or  
RNA-encoding) are now designated as gene synonyms instead of  
labels. This should allow them to be searched for in Entrez as gene  
names.

## FEATURES

source  
location/Qualifiers  
1..13176  
/organism="Escherichia coli K12"  
/mol\_type="genomic DNA"  
/strain="K12"  
/db\_xref="taxon:83333"  
/db\_xref="MG1655"  
/complement (49..81)  
/note="factor Sigma70; promoter hns; documented +1 at  
263862"  
complement (86..1204)  
/gene="gcpB"  
/note="synonym: b2515"  
complement (86..1204)  
/gene="gcpB"  
/function="orf; unknown"  
/note="f372: 100 pct identical to GCPE\_ECOLI SW: P27433"  
/codon\_start=1  
/transl\_table=11  
/product="orf, hypothetical protein"  
/protein\_id="AAC75568.1"  
/db\_xref="GI:1788863"  
/translation="MHNOAPRIORRSRTIYGVNPIGDCAPAIYAVOSMTNTRTIVENT  
VNOIKALERVGADIVRSVPTMDAAEFILIQOVNVPVADHFPYRIALAKAEYGV  
DCIRINPGRINGBERIRNWDCAKDNIPRIRGVNAGSLEKDIQERTEPTPALLES  
ARKHVDHLRLNFDQKVSADVPLAVESYRLAKQIDPHLIGITEGPGASGAV  
KSAIGLGLSEIGIDTLRVSADPVEIKYGFILKSLIRSRGINFIACPTCSRO  
EPDVGTVNALSGRLDITTPMDVSIIGCVNPGPALVSTIGTGKSKSGLYEDGV  
RKDRLDNMIDOLEARIKAKASQDEARIDVQVEK"  
complement (1231..2244)  
/gene="yfgA"  
/note="synonym: b2516"  
complement (1231..2244)  
/gene="yfgA"  
/function="putative membrane; Not classified"  
/note="f337; 100 pct identical to YFGA\_ECOLI SW: P27434"  
/codon\_start=1  
/transl\_table=11  
/product="putative membrane protein"  
/protein\_id="AAC75569.1"  
/db\_xref="GI:1788864"  
/translation="MNTERTHDQNEALTTCARLRNARBQGLSGQAVARLCLKVSTV  
RIEDKAPADLASTFLRGYIRSYARLVHPEELPGLEKQAPLEAAVAPQSFSL  
GRRKRKDDMLMTFWLVLFVVIIGSGAMWDRKQOEITTMQOASAELESNSQV  
GGSVPINTSTTDPAITSTPASVDVATNTQTPAYTAAPAVDPQNAVSPSQNV  
DPAATPAATATTPDGAAPLPDQAGVTPVADPNVLAWNFTADCELYTDAGKGLF  
SGMKQKDGMLNTGQAPYTKTGAPAAVIOYQKRVDSLRFRTNQVARTLTMAQDS  
PAQ"  
complement (1288..1316)  
/gene="yfgA"  
/note="factor Sigma70; predicted +1 start at 2639901"  
complement (1360..1389)  
/gene="yfgA"  
/note="factor Sigma70; predicted +1 start at 2639973"  
complement (2529..3683)  
/gene="yfgB"  
/note="synonym: b2517"

gene  
promoter  
promoter  
gene

misc\_feature  
4397..9255  
/note="O-island #106: Region of the EDJ93 chromosome not homologous to E. coli K-12 MG1655"  
gene  
complement(4405..4983)  
/gene="Z3782"  
complement(4405..4983)  
/gene="Z3782"  
/function="orf: Unknown function"  
/note="No significant matches"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="AAG57629.1"  
/db\_xref="GI:12516913"  
/translation="MLMKKGGITATAIDISQAKWERRALLNECLROYGEGWV  
SPHSEVNVSRRLPHYQADVOMTOIQDPLRLHILIPISAPRPMQKOLCGA  
CMFACEQVPSLENGHQLINDALCNGCONCIAECFOAMVELLILPAKIVLHNRK  
VCTCKQSFLLFOQNMNCLYQCRHRYGMRTP"  
CDS  
complement(5259..6080)  
/gene="Z3783"  
complement(5259..6080)  
/gene="Z3783"  
/function="putative enzyme; Not classified"  
/note="Residues 2 to 272 of 273 are 25.26 pct identical to residues 6 to 276 of 286 from GenPept 118 : gi15002126|gb|AAD37317.1|AF13170.8 (Af13170) dimethyl sulfoxide reductase subunit C [Yersinia pestis]"  
/codon\_start=1  
/transl\_table=11  
/product="putative dimethyl sulfoxide reductase subunit C"  
/protein\_id="AAG57630.1"  
/db\_xref="GI:12516914"  
/translation="MHELRPMAPLILQASVGMTLMARHRIILTSRPMRPL  
LAPLAGVGLISTHLGYPILNMAHLSFSSMSREITFEALYALGLSLVW  
LRGGWQLMLAAVGIADVFCMAQVYHTSIIITQHYTLIMFGVIGILSLVA  
VFRISGILPQIDALRNGCVIALVLRVLPWIGDITANMGIATLPHAPLML  
GQKPLITLSWGSVIGMFFAVGCKKNIPALFGSVMLVGEVNLRFVFSIG"  
CDS  
complement(6073..6510)  
/gene="Z3784"  
complement(6073..6510)  
/gene="Z3784"  
/function="putative enzyme; Not classified"  
/note="Residues 4 to 134 of 145 are 61.83 pct identical to residues 57 to 186 of 205 from GenPept 118 : gi11787122|gb|AAC73981.1|AB00191|anaerobic dimethyl sulfoxide reductase subunit B [Escherichia coli]"  
/codon\_start=1  
/transl\_table=11  
/product="putative dimethyl sulfoxide reductase subunit B"  
/protein\_id="AAG57631.1"  
/db\_xref="GI:12516915"  
/translation="MHNHFAVYTLVSCHNCADPCTKNCPTAMHRRPDDGIVRVNT  
DKVGGCYGAMSCPYGAPQNNQOTGMKSKDPCIDIAKQEQDICYATCELGAIKFGP  
IDLRKRYGVSIVKGLPDSITQPNLVIKPHOGAEKASHHA"  
CDS  
complement(6699..9080)  
/gene="Z3785"  
complement(6699..9080)  
/gene="Z3785"  
/function="putative enzyme; Not classified"  
/note="Residues 25 to 792 of 793 are 43.59 pct identical to residues 3 to 784 of 785 from GenPept 118 : gi11787121|gb|AAC73980.1|AB00191|anaerobic dimethyl sulfoxide reductase subunit A [Escherichia coli]"  
/codon\_start=1  
/transl\_table=11  
/product="putative dimethyl sulfoxide reductase subunit A"  
/protein\_id="AAG57632.1"  
/db\_xref="GI:12516916"  
/translation="MDPPDKERITPPQVSRSLQATSLITLPIRISSTAKQSDA  
SEPTAPVADKVPPTCSFGCGKCDIRAMRGVATQITLIDNEDLPMPIMRACV  
RGGRYKRVYHPRLKYPKRGKGRKSERISWDEATLIDNKRITLQYQSPSR  
YVAVGTAIVGGTSGDAMARLLNLTGQYLEYVHSLGNTAATATYTGVAAGSM  
DTLDTKLVIIMGNPTETIFGHTNYFQMKONGTRFIIVDBRYSVTSADQMP

Alignment Scores:  
Pred. No.: 1,98e-134 Length: 11521  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
US-09-921-992-50 (1-372) x AB005481 (1-11521)  
QY 1 MethiAsnGluAlaProIleGlnArgGlySerThrArgIleTyrValGlyAsnVal 20  
DB 1204 ATGCATACCGAGGCTCCAAATTCACGTAAGAAATCAACGTAATTCATTGGGAAATGIG 1145  
QY 21 ProIleGlyAspGlyValProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
DB 1144 CGGATTGGGAGTGTCTCCCATCGCGCTGACATGACATGACCAATACCGGTACGACAGAT 1085  
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60  
DB 1084 GTCAAGCAACAGGTCATCAATCAAGCGCGTGGACGCGTGGCGCTGATATCGTCCGT 1025  
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80  
DB 1024 GTCTCCGTCGACGATGAGCGGAGGAGGAGGCTTAACACTCAACAAACGCGATTAC 965  
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
DB 964 GTCCCGCTGTGGCTGACATCCACTTCGACTATCGCATTCGCGTGAAGTAGCGGATAC 905  
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluLysIleArgMet 120  
DB 904 GCGGTGATTTGCTGCTGATTTAACTTCGCAATATCGTAATGAMAGCGTATTCGCATG 845  
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
DB 844 GTGGTGTGACGTGGCGGCGCAATAAACATTCGATCGATATGGCCGTAAACGCGGATCG 785  
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
DB 784 CTGGAAAGAAATCTGCAGAAAGAAAGTATGGCGAAGCAGCCCGACGCGTTCGATCC 725  
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
DB 724 GCATCGCGCATGTTGATCATCTCGATCCCTGMACTTGATGATCAAGTCAACGCGTG 665  
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
DB 664 AAGCGCTGACGCTCTCTCGTGTGAGCTTATCGTTCGCAAAACAAATCGAT 605  
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220  
DB 604 CAGCGCTCATCTGGGATCAACGAGGCGGCGCGCGACGCGGAGGAGTAAATCC 545  
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
DB 544 GCATTTGTTTGTCTGCTGCTGCTGATGAAAGGATGCGGACACGCTGCGCATCATCG 485  
QY 241 AlaAlaAspProValGluGluLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
DB 484 GCGGCGGATCGGTCGAAAGATCAAGTGGTTTGATATTTGAAATCGCTTCATAC 425  
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280

QY	261	11gcrsrraaggyllleaasrphellleaacyserprothrCysseraaggingluPhaaspva1	280
Db	1315	CGTTCGCGAGGAGATCAACTTCATCCCTCGCCCGCACCTGTTCCGCGCAGAAATTTGATGTT	13744
QY	281	11eglythrValaAsna1alaLeug1ug1nargrleug1uaAsp1le1ethProMetAspva1	300
Db	1375	ATCGGTACGTTAAACGGCTGGAGCAACGCTGGAAAGATCAATCACTCCATGGACCTT	1434
QY	301	Ser1le1leglyCyseValaValaAsnGlyProGlyGua1aleuValaSerThrLeuGlyVala1	320
Db	1435	TCGATTAATCGCCTGCTGCTGTAATGCCCCAGGTGAGCGCGCTTTCTACACTCGGCGTC	1494
QY	321	ThrglyGlyAanlyValySerGlyVleuThrGlyGua1aspGlyValaLyGlyAspArgLeuasp	340
Db	1495	ACCGCGGCAACAGAAAGAGCGGCTCTATGAAGATGGCGGCGCAAGACCGTCTGGAC	1554
QY	341	AasnAspMet1leaSpGlnLeuGlnua1aarg1leaGlyAlaLyAsaSerGlnLeuasp	360
Db	1555	AACAACGATATGATGATGACCACTGGAAACACGCAATTCGTGGAAAGCCAGTCACTGGAC	1614
QY	361	GLua1aargyglleaSpValaGlnGlnValaGluLyS	372
Db	1615	GAAGCGCTCGAATTGACGTTCAAGAGGTTGAAAAA	1650
RESULT 7			
AE005481/c	AE005481	11521 bp	DNA linear BCT 21-MAR-2001
LOCUS	Escherichia coli O157:H7 EDL933	genome, contig 3 of 3,	section 100
DEFINITION	of 290.		
ACCESSION	AE005481	AE005174	
VERSION	AE005481.1	GI:12516908	
KEYWORDS			
SOURCE			
ORGANISM	Escherichia coli O157:H7 EDL933		
	Escherichia coli O157:H7 EDL933		
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
	Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 11521)		
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D.,		
	Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,		
	Pofofai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,		
	Grobeck,E.U., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,		
	Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,		
	Weich,R.A., and Blattner,F.R.		
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7		
JOURNAL	Nature 409 (6819), 529-533 (2001)		
MEDLINE	21074935		
PUBMED	11206551		
REFERENCE	2 (bases 1 to 11521)		
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D.,		
	Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,		
	Pofofai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,		
	Grobeck,E.U., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,		
	Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,		
	Weich,R.A., and Blattner,F.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of		
	Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		
FEATURES			
Source	Location/Qualifiers		
	1..11521		
	/organism="Escherichia coli O157:H7 EDL933"		
	/mol_type="genomic DNA"		
	/strain="EDL933"		
	/serotype="O157:H7"		
	/db_xref="taxon:155864"		
	/note="enterohemorrhagic"		
	complement(86..1204)		
	/gene="gcpB"		
	/note="synonym: 23778"		
	complement(86..1204)		
	/gene="gcpB"		
	/function="orf; Unknown function"		
	/note="Residues 1 to 372 of 372 are 100.00 pct identical		

to residues 1 to 372 of 372 from *Escherichia coli* K-12

Strain MG1655: B2515"

    /codon\_start=1

    /transl\_table=11

    /product="orf, hypothetical protein"

    /protein\_id="AAG57625.1"

    /db\_xref="GI:12516909"

    /translation="MNNAKPIQRKSTRIVGVNGVIGDAPAVIVDSMTNRTDVEATVNOIKALEVGVADIVRVSPIPTMDAAKAEFLIKQAVNPLVADIDHIDRIALKAIEYVDCILRINPNGIINGEERIRMWVDCARDNIIPIRIGVNGSLSEKDIQEKYEPTPALLESSARAHVDLRLNLEDPQKSVKASDVEIAVESYELAKQIDPHLGITTEGARGSGVAKSAIGGLGLSLGIGIPKASDAVEEIKVGFILKSLRISRGINFIACPSCSQEDVDGIVTVALSQRLEDITITPMDVSIIGCVNNGPGLVSTLGVTSNKKSLGYEDGVKRDRLDNNMTIQLEKIRAKISQDLDEARRIDVQVEK"

    complement(1231..2244)

    /genes="yFga"

    /feature="putative membrane, Not classified"

    /note="Residues 1 to 337 of 337 are 99.10 pct identical to residues 1 to 337 of 337 from *Escherichia coli* K-12 strain MG1655: B2516"

    /codon\_start=1

    /transl\_table=11

    /product="putative membrane protein"

    /protein\_id="AAG57626.1"

    /db\_xref="GI:12516910"

    /translation="NMYEATHQONEALTGGARLRNAREQLGSLQAVAEELCLKVSTVRIIEEDKADPLASITFLVIRISYVARLVIPEEELPLGEKQALPAKVAAPMOSFELGRRKKRQPMATLPTMLVLPVITIGLSGAMWMDHKQOEIITMADOSSKEISNSPQGVSVLNTXTTTPDATTSTPPASVDTTATNTQTPVTAAPAVDPQONAVSPSQAVDTPQAPAPAPATPAPGAALPLDQAGVTPVADPNALVNNFTADCWLEVDADGKLFPSMAQPKDGLNLTNGGAPVPLKLIQGAAPAAVOIQYQKPEVDLSFRTIQVALTLNAEOSPIQ"

    complement(2529..3683)

    /genes="yFgb"

    /note="synonym: Z3780"

    complement(2529..3683)

    /feature="orf: Unknown function"

    /note="Residues 1 to 384 of 384 are 99.47 pct identical to residues 1 to 384 of 384 from *Escherichia coli* K-12 Strain MG1655: B2517"

    /codon\_start=1

    /transl\_table=11

    /product="orf, hypothetical protein"

    /protein\_id="AAG57627.1"

    /db\_xref="GI:12516911"

    /translation="MSEQLVTEVNTTKDGKINLIDLNRQNEFFPKDIGEKPRADQVMKMYHYCCDNFDEMTDINKYRGKLKEVASIRAPVEVEQRSSDGTIKMAVAGQO RVEYTIIEEDDATTCAVSSQVCALEKRCSTAOQGFNNLRVSEIIGQVRAKIIQGLAKVAGORPITNVMMGGEPLINNNVPAEIMLDDRGFGLSKRVLITSGVPA LDKGDMIDVALAIIELHAPNDIRBEIIVINCKXNKTETLAARLRILESNNAQGRVTLSYGLVLDHNDSTGRHDAQAEILKQTPCKINIIIPMPFPAAPGRSSNIDFBSKVLMSYGVTLTVARKTGDDIDIAACAGDVIDRTKRTLKKMGGEALIDIKAV"

    complement(3833..4264)

    /genes="ndk"

    /note="synonym: Z3781"

    complement(3833..4264)

    /genes="ndk"

    /function="enzyme; Nucleotide biosynthesis: Purine ribonucleotide biosynthesis"

    /note="Residues 1 to 143 of 143 are 100.00 pct identical to residues 1 to 143 of 143 from *Escherichia coli* K-12 Strain MG1655: B2518"

    /codon\_start=1

    /transl\_table=11

    /product="nucleoside diphosphate kinase"

    /protein\_id="AAG57628.1"

    /db\_xref="GI:12516912"

    /translation="MAIERTSIIIPKNAVAKVIGNIPARFEAGFKIVGTKMLHLTVIGARFVAEHOKPEPDDGLVEMTSGPIVSVLEGENAVQARRHDLGATNPAAAGT

Qy 261 ArgSerArgIyleaenphei1eAaCyProthrcCySerArgIngluPheaspVal 280  
Db 781 CgTTCGGAGGATCACTTCATCGCTGCCACCTGTCGCTCGAGAAATTGATGTT 840  
Qy 281 IlGlyThrValaenalaLeugluInArgLeugluAspIleIleThrPmetAspVal 300  
Db 841 ATCGGATACGGTTAAACGGCTGAGCAACGCTGGAAAGATATCATCATCCGATGAGACTT 900  
Qy 301 SerIleIleGlyCySerValaLeugluProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 901 TCGATTATCGCTGCGCTGTAATGCGCCAGGTGAGCGCTGTTCTTACACTCGGCGTC 960  
Qy 321 ThGlyGlyAsnIlySerGlyLeuTyGlyAspGlyValaArgLyAspArgLeuAsp 340  
Db 961 ACCGCGCGCAACAGAAACCGGCTTATGAAAGATGCGCTGCGCAAGACCGTCTGAGC 1020  
Qy 341 AsnAsnAspMetIleAspGluLeuGluAlaArgIleArgAlaLyAspArgInLeuAsp 360  
Db 1021 AACACGATATGATCGACCGAGTGAAGACGCGATTGTCGGAAGCCAGTCACTGAGAC 1080  
Qy 361 GluAlaArgArgIleAspValaGluIngluValaGluLys 372  
Db 1081 GAAGCGCGTGAATGACGTTACAGCAGTTGAAAAA 1116  
RESULT 6  
ECGCE 1697 bp DNA linear BCT 08-DEC-1993  
LOCUS E.coli gcpe gene. 1697 bp DNA linear BCT 08-DEC-1993  
DEFINITION X64451 S43432  
ACCESSION X64451.1 GI:41540  
VERSION 6  
KEYWORDS gcpe gene.  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
1 (bases 1 to 1697)  
Parker, J.  
Direct Submission  
Submitted (17-FEB-1992) J. Parker, Southern Illinois Univ., Dept of  
Microbiology, S I U, Carbondale IL 62901, USA  
2 (bases 1 to 1697)  
Baker, J., Franklin, D. B. and Parker, J.  
Sequence and characterization of the gcpe gene of Escherichia coli  
FEMS Microbiol. Lett. 94, 175-180 (1992).  
See also J01629 & M11843.  
FEATURES  
Location/Qualifiers  
1..1697  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/strain="K12"  
/db\_xref="taxon:562"  
/sub\_clone="PSE411"  
/cell\_line="JF238"  
/clone\_lib="lambda transducing phage dguab-2"  
1..508  
/note="ORF"  
/codon\_start=2  
/transl\_table=11  
/protein\_id="CAA45782.1"  
/db\_xref="GI:41541"  
/db\_xref="SWISS-PROT:P27434"  
/translation="TTTDPATMTTPASVDTTATNTCTAPAVTAPAPAVPQCNAAVSP  
SQANVTATPATPATATTPDGAAPLPDAGCTTTPVADPNALVMTADCMLEVTAT  
GKLFSGMQRKDNLMVTGAPYKLTIGAPAAVIOYQGPVLSRFIRTNQVRLTL  
NAQSPAQ"  
535..1653  
/gene="gcpe"  
535..1653  
/gene="gcpe"  
/codon\_start=1  
/transl\_table=11  
/protein\_id="CAA45783.1"

/db\_xref="GI:41542"  
/db\_xref="SWISS-PROT:P27433"  
/translation="MHNQAPLQRRKSTRVGNVPIGDPAPLAVNQMTRTTPDVAT  
VNOIKALERVGDYIRVSPPTMDAAEFLLIKQNVNVLADIHPIYRALKVAEYGV  
DCLRINPAGINBERIRNVDCARDKNIPIRIGVNAQSLEKDLQERYGEPYQALLES  
ARRVHDHLDRLNEFDQKVSVDFAKNIIRKVAESRLAKQIDQPHLQITAGARSGAV  
KSAIGIGLLISEGIDGIFDLKSLADPVEEIKGFDLKSLIRSRINFIACPTCSRO  
EDVGIVTVALORLEDDITTPMDVSIIGCVNNGPGLAVSTIGVTGNKSGLYEDGV  
RDRPLDNDMDQLERIRAKSOLDEARIDVQVEK"  
BASE COUNT 417 a 446 c 469 g 365 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,58e-135 Length: 1697  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 1 Gaps: 0  
US-09-921-992-50 (1-372) x ECGCE (1-1697)  
Qy 1 MethIeAsnGluAlaProIleGluArgLySerThrArgIleTyValaGlyaenVal 20  
Db 535 ATGCATTAACCAAGCTCCCAATTCACAGTAAATCAACAGTATTACGTGGGAATGTG 594  
Qy 21 ProlIleGlyAspGlyValaProIleAlaValaGluInserMetThrAsnThrArgThrAsp 40  
Db 595 CCGATTGGCGATGTCCTCCATCGCTGACATCATACCAATACGATGAGTACGACAGAC 654  
Qy 41 ValGluAlaThrValaAsnGluInleuValaLeuGluArgValaGlyValaAspIleValaArg 60  
Db 655 GTCGAAGCAACGGTCATCAATCAAGCGCGTGAAGCGGTTGCCCTATATCGTCCGT 714  
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGluIngluValaen 80  
Db 715 GATACCGTACCGACGATGAGCGCGGCAAGCGTTCAACTCATCAACAGCAGGTTAAC 774  
Qy 81 ValProLeuValAlaAspIleHisPheAspTyArgIleAlaLeuLysValaGluTyT 100  
Db 775 GGCCCGCTGGTGGCTGACATTCACCTTCGATTCGCTGAAAGTACGGGAATAC 834  
Qy 101 GlyValaAspCyLeuAspArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
Db 835 GCGGTGATGTCGTGCTGATTAACCTGCAATATGTAAGAGCGTATTCGATG 894  
Qy 121 ValValaAspCyAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140  
Db 895 GTGGTTGACTGTGCGGCGCATTAACATTCGATTCGATTGGCGTTAACCGCGATCG 954  
Qy 141 LeuGluLysAspLeuGluIngluLysTyArgIluProThrProGluAlaLeuLeuGluSer 160  
Db 955 CTGGAATAAAGATCTGCAGAAAGATATGCGAACCACCCGCGCGTTCGTGATCT 1014  
Qy 161 AlaMetArgHisValaAspHisLeuAspArgLyAsnAsnPheAspGluPheLysValSerVal 180  
Db 1015 GCCATCGCTCATGTTGATCATCTCGATCCCTGCACTTGATGATCAAGTACAGCGTG 1074  
Qy 181 LysAlaSerAspValaPheLeuAlaValaGluSerTyArgLeuLeuAlaLysGluInleAsp 200  
Db 1075 AAAGCGTTCAGCGTCTTCCTCGCTGTGAGTCTTATCGTTTCTGCAAAACGATCGAT 1134  
Qy 201 GlnProLeuHisLeuGluGlyIleThrGluAlaGlyValaLysSerGlyValaLysSer 220  
Db 1135 CAGCGCTTCATCTGGGATCAACGAAAGCGGTGGCGCGACGCGGCGAGTAAATCC 1194  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
Db 1195 GCCATTGTTAGTGTCTGCTGCTGTAAGCATCGGAGACACGTCGCGCTATCGCTG 1254  
Qy 241 AlaAlaAspProValaGluGluIleLysValaGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 1255 GGGGCGGATCCGGTGAAGAGATCAAAAGTGGTTCGATATTTTGAAATCGCTGCGTATC 1314



Db 421 CTGGAAGAAAGATCTGCAAGAAAGTATGGCAACCGACCGCGAGCGGTTCGTGATCT 480  
Qy 161 AAlaMetArgHisValAspHisLeuAspArgLeuAspPheAspGlnPheLeuValSerVal 180  
Db 481 GCCATGGGTATGATGATCTGATCGCTGATCGCTGATCGATCGATCGATCGATCGATCG 540  
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
Db 541 AAGCGCTGACGCTCTTCCTCGCTGTTGATCTTATCGTTTGGTGGGAAAACGATCGAT 600  
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyIleValAlaArgSerGlyAlaValLysSer 220  
Db 601 CAGCCGCTGATCGGGGATCACCAGAACCGGTGTGCGCGACGCGGGCAGTAAATCC 660  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyIleValPheLeuValSerLeu 240  
Db 661 GCCATTGGTTAGTGTCTGCTGCTGTAAGAGCATCGGACACGCTGCGATCGCTG 720  
Qy 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 721 GCGGCGGATCGGTCGAAAGATCAAAAGTCGTTTCGATTTTGAATTCGCTCGATC 780  
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280  
Db 781 CGTTCGCGAGGATCAAACTTCATCGCTGCGCCGACCTGTCGCTGAGAAATTTGATGTT 840  
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300  
Db 841 ATCGGATCGGTTAAACGCGTGGAGCAACGCTGGAAGATTCATCATCTCGATGGACGTT 900  
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 901 TCATATATCGGCTCGCTGCGTGAATGGCCAGTGAAGCGCTGTTTCTAATCTCGGCTC 960  
Qy 321 ThrGlyIleValAsnLysSerGlyLeuTyrGlyIleAspGlyValAlaArgLysAspArgLeuAsp 340  
Db 961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGCGCTGCGCAAGACCGCTCGAC 1020  
Qy 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAspArgLeuAsp 360  
Db 1021 AACAAAGATATGATTCACACGCTGGAAGACGCAATTCGTGCGAAAGCCATCGCTGAC 1080  
Qy 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372  
Db 1081 GAAGCGCTGATTAATTAAGCTTCAGACGCTTGAAGAAA 1116

RESULT 5  
AX393943 1119 bp DNA linear PAT 23-MAR-2002  
LOCUS Sequence 3 from Patent WO0212478.  
DEFINITION AX393943  
ACCESSION AX393943  
VERSION AX393943.1 GI:19701905  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE  
AUTHORS Seeman, A., Campos, N., Rodriguez-Concepcion, M., Rohmer, M.,  
Serrano, M., Valentin, H.E., Venkatesh, T.V. and Venkatesh, M.  
TITLE Methyl-d-erythritol phosphate pathway genes  
JOURNAL Patent: WO 0212478-A 3 14-FEB-2002;  
Monanto Technology LLC (US)  
FEATURES  
source Location/Qualifiers  
1..1119  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
1..1119  
/note="unnamed protein product"  
/codon\_start=1  
/transl\_table=11

/protein\_id="CAD29065.1"  
/db\_xref="GI:19701906"  
/translation="MKNQAPIQRRKSTRIVGVNPIGDGAPLAVQSMNTNRTTVEAT  
VNOIRALERVQADIRIVRVSPTMDAAEFLTIQOVNVRVAIHPFYRLAKVAREGV  
DCLRINPQIGNBERIRKRVYDCAKDNIPIRIGLVNAGSEKDIQKRGTEPTPALLS  
ARRHVDHLRLNFDQKVSVKASDVFLAESTRLKQIDQPLHKTITAGAGRSAGV  
KSAIGLGLSBSGIDGTLRVSADPVEIKGFDLKLRSRGINIACPTCSRQ  
EPDVGITVALQRLIEDITTPMDVSGIGCVNAGPGBALVSTLGTGANKSGLYEDGV  
RDRLDNNMIDQLEARIKAKASQDEARIDVQVEK"

BASE COUNT 272 a 279 c 317 g 251 t  
ORIGIN

Alignment Scores:  
Pred. No.: 9,1e-136 Length: 1119  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: Gaps: 0

US-09-921-992-50 (1-372) x AX393943 (1-1119)

Qy 1 MethIleAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20  
Db 1 ATGCATTAACCAAGCTTCAATTCACAGTAAATCAACGTAATTAACGTTGGAAATG 60  
Qy 21 ProIleGlyAspGlyValaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
Db 61 CGGATTGGGATGAGTCTCCATCGCTGACGTCCATACCAATACGATGAGGATGAC 120  
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyValaAspIleValArg 60  
Db 121 GTCGAAGCAACGGTCAATCAATCAAGCGCTGGAACGCTGGCCGTATATCGTCCG 180  
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80  
Db 181 GATATCCGTACCAACGATGACCGCGCAGAGGCTTCAAACTCATCAACAGCAGTTAAC 240  
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValaGluTyr 100  
Db 241 GTGCGCTGTGTGCTGACATTCACCTTCGATTCGATTCGCTGAAAGTACGGAATAC 300  
Qy 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
Db 301 GCGCTGATGCTGCTGCTATTAACCTGSCATATCGTAATGAAGAGCTATCCCATG 360  
Qy 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140  
Db 361 GTGCTTGAATGTCGCGCGATTAACCATTCGATCCGATTCGATTAACCGCGATCG 420  
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyIleProThrProGlnAlaLeuLeuGlnSer 160  
Db 421 CTGGAAGAAAGATCTGCAAGAAAGATGCGCAACCAACCGCCGAGCGCTGCGAATCT 480  
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLeuValSerVal 180  
Db 481 GCCATGCGCATGTTGATCATCTCGATCGCTGAACTTGATCAAGTACACGCTG 540  
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
Db 541 AAGCGCTGACGCTCTTCCTCGCTGTTGATCTTATCGTTTGGTGGGAAAACGATCGAT 600  
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyIleValAlaArgSerGlyAlaValLysSer 220  
Db 601 CAGCCGCTGATCGGGGATCACCAGAACCGGTGTGCGCGACGCGGGCAGTAAATCC 660  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyIleValPheLeuValSerLeu 240  
Db 661 GCCATTGGTTAGTGTCTGCTGCTGTAAGAGCATCGGACACGCTGCGATCGCTG 720  
Qy 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 721 GCGGCGGATCGGTCGAAAGATCAAAAGTCGTTTCGATTTTGAATTCGCTCGATC 780



QY 41 ValGluAlaThrValaLengInlelValaLengInuArgValaGlyAlaAspIleValArg 60  
 Db 121 GTGAAACCAACGGTCAATCAATCAAGCGCTGGAAGCGCTGGCGTGAATCGTCCGT 180  
 QY 61 ValSerValProThrmecAspAlaAlaGluAlaPheLysLeuIleLysGInGInValaSn 80  
 Db 181 GTATCCGTACCGACGATGAGCGCGCAAGACGTTCAACTCAATCAACACGAGTTTAC 240  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlyTyr 100  
 Db 241 GTGCGCGTGGTGGTGCATCTCACTCGACTATGCGATTGGCGTGAAGTAGCGGAATAC 300  
 QY 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyLysGInGInuArgIleArgMet 120  
 Db 301 GGGCGTCAATGTCTGCTATTAACTCGCAATATCGGTATGAAGAGCGTATTCGCATG 360  
 QY 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAlaAspIleSer 140  
 Db 361 GTGCTTACCTGTGGCGCGCATTAACATTCGATCCGATTCGGCTTAACGCCGATCG 420  
 QY 141 LeuGluLysAspLeuGInGInuLysTyrGlyGluProThrProGInAlaLeuLeuGlySer 160  
 Db 421 CTGGAATAAAGATCTGCAAGAAAGTATGGCGAACCGACCGCGCGCTTGTCTGGATCT 480  
 QY 161 AlalMetArgHisValaAspHisLeuAspArgLeuAsnPhaAspGInPhelysValSerVal 180  
 Db 481 GCCATGGGTATGTATGATCTGATCGCTGATCGCTGAATCTCAATCAAGTCAAGCGTGG 540  
 QY 181 LysAlaSerAspValaPheLeuAlaValaGlySerTyrArgLeuLeuAlaLysGInIleAsp 200  
 Db 541 AAGCGCTGACGCTTCTCTCGCTGTGATGCTTATGCTTTGCTGGCAAAACAGATGAT 600  
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220  
 Db 601 CAGCGCTGATCTGGGGATCAACGAAGCGGTGTCGCGCGACGCGGAGTAAATCC 660  
 QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
 Db 661 GCCATTGGTTAAGTGTGCTGCTGTCTGAAGCAATCGCGCACGCGTGGCTATCCCTG 720  
 QY 241 AlaAlaAspProValaGluGluIleLysValaGlyPheAspIleLeuLysSerLeuArgIle 260  
 Db 721 GCGGCCATCCGCTCGAAGAGATCAAACTCGCTTCAATTTTGAATCCGCTCGATC 780  
 QY 261 ArgSerArgGlyIleAsnPhaIleAlaCysProThrCysSerArgGInuPheAspVal 280  
 Db 781 CGTTCGAGGAGATCACTTCATCGCTGCGCGACCTGTCGCTCAGAAATTTGATGTT 840  
 QY 281 IleGlyThrValaAlaLeuGluGInArgLeuGluAspIleIleThrProMetAspVal 300  
 Db 841 ATCGGTACGGTTAAGCGCTGAGCAACGCTGGAAGATATCATCATCCGATGAGCGTT 900  
 QY 301 SerIleIleGlyCysValaValaAsnGlyProGlyGInuAlaLeuValaSerThrLeuGlyVal 320  
 Db 901 TCGATTATCGCTCGTGTGTAATGCGCCAGGTAGCGCGCTGTTTCTACCTCGCGCTC 960  
 QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValaArgLysAspArgLeuAsp 340  
 Db 961 ACCGCGGCAACAGAAAGCGCGCTCATATGAATGAGCGCGCAAAAGCCGCTGAGAC 1020  
 QY 341 AsnAsnAspMetIleAspGInLeuGluAlaArgIleArgAlaLysAlaSerGInLeuAsp 360  
 Db 1021 AACACACATATGATGACACACTGGAAGCAGCATGTGTGGAAGCAGTCACTGAC 1080  
 QY 361 GlnAlaArgArgIleAspValGInGInuValaGlyLys 372  
 Db 1081 GAAGCGGTGCAATTGACGTTCAAGAGTTGAAAAA 1116  
 RESULT 4  
 AX050487 1119 bp DNA linear PAT 12-JAN-2001  
 LOCUS AX050487  
 DEFINITION Sequence 1 from Patent WO0072022.

ACCESSION AX050487  
 VERSION AX050487.1 GI:12226691  
 KEYWORDS  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
 REFERENCE  
 1 Jomaa, H.  
 TITLE Use of genes of the deoxy-d-xyulose phosphate biosynthetic pathway for altering the concentration of isoprenoid  
 JOURNAL Patent: WO 0072022-A 1 30-NOV-2000;  
 Jomaa, Hassan (DB)  
 FEATURES  
 source Location/Qualifiers  
 1..1119  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 1..1119  
 /note="unnamed protein product"  
 /codon\_start=1  
 /transl\_table=11  
 /protein\_id="CAC21801.1"  
 /db\_xref="GI:12226692"  
 /translation="MHNOAPIORRKRSTRIVGVNPFGDGPAPVAVQSMTRTRTDEAT VNOIKALERVADIVRVSVPTMDAARAFILIQOVNVPVADHFPDRIALAKVAEYGV DCLRIINPNIENBERIRMYDDCARDNIPIRIGVNAKQIDQPLHGIITEGARKGAV AMKHVDHDLRNPDPKVSVDVFLAVESIRLVAKQIDQPLHGIITEGARKGAV KSAIGLGLLSGIGDILRVSLAADPVEISIKVGFILKSLRISRGINFLACPTCSRO EFDVIGTVALRQLEDDIITPMDVSIIGCVNPGFALVSTLGVTKNKSGLYEDGV RDRLDNNMIDOLEARIRAKASQIDEARIDVQVEK"  
 BASE COUNT 272 a 279 c 317 g 251 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9..1e-136 Length: 1119  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: Gaps: 0  
 US-09-921-992-50 (1-372) x AX050487 (1-1119)  
 QY 1 MethiAsnGInAlaProIleGInaArgGlySerThrArgIleTyrValaGlyAsnVal 20  
 Db 1 ATGCATTAACCAAGCTCAATCAACGTAGAAATCAACAGTATTAACCTTGGAAATGTG 60  
 QY 21 ProIleGlyAspGlyAlaProIleAlaValaGlnSerMetThrAsnThrArgThrAsp 40  
 Db 61 CCGATTGGCGATGTGCTCCCATCGCGCTGACATCCATGACCAATACCGCTACGACAGAC 120  
 QY 41 ValGluAlaThrValaAsnGInuLysAlaLeuGluArgValaGlyAlaAspIleValArg 60  
 Db 121 GTGAAACCAACGGTCAATCAATCAAGCGCTGGAAGCGCTGGCGTGAATCGTCCGT 180  
 QY 61 ValSerValProThrmecAspAlaAlaGluAlaPheLysLeuIleLysGInGInValaSn 80  
 Db 181 GTATCCGTACCGACGATGAGCGCGCAAGACGTTCAACTCAATCAACACGAGTTTAC 240  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlyTyr 100  
 Db 241 GTGCGCGTGGTGGTGCATCTCACTCGACTATGCGATTGGCGTGAAGTAGCGGAATAC 300  
 QY 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyLysGInGInuArgIleArgMet 120  
 Db 301 GGGCGTCAATGTCTGCTATTAACTCGCAATATCGGTATGAAGAGCGTATTCGCATG 360  
 QY 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAlaAspIleSer 140  
 Db 361 GTGCTTACCTGTGGCGCGCATTAACATTCGATCCGATTCGGCTTAACGCCGATCG 420  
 QY 141 LeuGluLysAspLeuGInGInuLysTyrGlyGluProThrProGInAlaLeuLeuGlySer 160

VERSION AX036302.1 GI:11225912  
KEYWORDS Escherichia coli  
SOURCE Escherichia coli  
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1  
AUTHORS  
TITLE Novel method for identifying antibacterial compounds  
JOURNAL Patent: EP 1043403-A 29 11-OCT-2000;  
GPC AG GENOME PHARMACEUTICALS (DE)  
FEATURES  
source 1. 1119  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
BASE COUNT 272 a 279 c 317 g 251 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 9,1e-136 Length: 1119  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-921-992-50 (1-372) x AX036302 (1-1119)  
QY 1 MethiAsnGlnAlaProIleGlnArgLysSerThrArgIleTyValGlyAsnVal 20  
DB 1 ATGCATACCAAGCTCCATTCAACGTAAGAAATCAACAGTATTACGTTGGGAATGTG 60  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40  
DB 61 CCATTGGCGATGGTGTCTCCATCGCGTACAGTCCATGACCAATACGCTACGACACGAC 120  
QY 41 ValGlnAlaThrValAsnGlnIleTyValAlaGlnArgValGlyAlaAspIleValArg 60  
DB 121 GTGCAAGCAACGGTCATCAATCAACGACCGCTGGACCGCTGATTCGTCCT 180  
QY 61 ValSerValProThrMetAspAlaAlaGlnAlaPheLysLeuIleLysGlnGlnValAsn 80  
DB 181 GTATCCGTCACGACGATGACGCGCGGAGGCGTTCAACTCAATCAACAGCAGGTTAAC 240  
QY 81 ValProLeuValAlaAspIleHisPheAspTyThrArgIleAlaLeuLysValAlaGlnTyr 100  
DB 241 GTGCCGCTGGTGGCTGCATCCACTTCGACTATCGCATTCGCGTGAAGAGCGGATAC 300  
QY 101 G1ValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnGlnArgIleArgMet 120  
DB 301 GGGCTGATGTCTGCGTATTAACCTGCGCATATCGGTAAATGAAGCGTATTCGATG 360  
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140  
DB 361 GTGGTTCATCTGTGGCGCGCATTAACATTCGATTCGATTCGCGTTAAGCGGATCG 420  
QY 141 LeuGlnLysAspLeuGlnGlnLysLysLysLysLysLysLysLysLysLysLysLys 160  
DB 421 CTGGAAGAAAGATCTGCAGAAAGATAGCGAAGCGGACGCGGCGGTGCTGGATCT 480  
QY 161 AlMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
DB 481 GCCATGGCTATGTTGATCTCTCGATCGCTGATCGATTCGATTCGATTCGATTCG 540  
QY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyThrArgLeuAlaLysGlnIleAsp 200  
DB 541 AAGCGCTGACGCTCTCTCGCTGCTGATCTTATCTGCTGGGAAACGATGAT 600  
QY 201 GlnProLeuHisLeuGlnIleThrGlnAlaGlyAlaArgSerGlyAlaValLysSer 220  
DB 601 CAGCCGTTGATCTGGGATCACCGAAGCGGCTGTCGCGCAGCGGCGGAGTAAATCC 660

QY 221 AlaIleGlyLeuGlnLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeu 240  
DB 661 GCCATTGGTTAGCTGTGCTGCTGTGAAGCATCGGCGACACGCTGCGGATCGCTG 720  
QY 241 AlaAlaAspProValGlnGlnIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
DB 721 GCGCCGATCCGGTCCAGAGATCAAGTCGATTTGATATTTGAAATCGCGCTATC 780  
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280  
DB 781 CATTCCGAGGATCAACTTCACTGCTGCTGCGACCTGCTGCGTCAAGAAATTTGATGTT 840  
QY 281 IleGlyThrValAsnAlaLeuGlnArgLeuGlnAspIleIleThrProMetAspVal 300  
DB 841 ATCGGTACGGTTAACCGCTGAGCAACCGCTGGAAGATATCATCTCGATGGAAGCTT 960  
QY 301 SerIleIleGlyCysValValAlaGlnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320  
DB 901 TCGATTATCGGCTGCTGCTGGAATGCGCCAGGTCGAGCGCTGCTTCTACACTCGCGCTC 960  
QY 321 ThrGlyValAsnLysSerGlyLeuTyThrGlnAspGlyValArgLysAspArgLeuAsp 340  
DB 961 ACCGCGCGCAACAGAAAGCGGCTCTTATGAAGATGCGCTGCGCAAGACGCTTGAC 1020  
QY 341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
DB 1021 AACCAACGATATGATTCACCGAGCTGGAACCGCATTCGTGCGAAGCCAGTCACTGAGAC 1080  
QY 361 GlnAlaArgArgIleAspValGlnGlnValGlnLys 372  
DB 1081 GAAGCGCTGCAATGACGTTGACGAGTGAAGAAA 1116  
RESULT 3  
LOCUS AX038912 1119 bp DNA linear PAT 16-NOV-2000  
DEFINITION Sequence 29 from Patent WO061793.  
ACCESSION AX038912  
VERSION AX038912.1 GI:11228221  
KEYWORDS  
SOURCE Escherichia coli  
ORGANISM Escherichia coli  
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
Enterobacteriaceae; Escherichia.  
REFERENCE 1  
AUTHORS Loferer H. and Jacobi A.  
TITLE Novel method for identifying antibacterial compounds  
JOURNAL Patent: WO 0061793-A 29 19-OCT-2000;  
GPC BIOTECH AG (DE) ; LOFERER HANNES (DE) ; JACOBI ALEXANDER (DE)  
FEATURES  
source 1. 1119  
/organism="Escherichia coli"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:562"  
BASE COUNT 272 a 279 c 317 g 251 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 9,1e-136 Length: 1119  
Score: 1866.00 Matches: 372  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
Gaps: 0  
US-09-921-992-50 (1-372) x AX038912 (1-1119)  
QY 1 MethiAsnGlnAlaProIleGlnArgLysSerThrArgIleTyValGlyAsnVal 20  
DB 1 ATGCATACCAAGCTCCATTCAACGTAAGAAATCAACAGTATTACGTTGGGAATGTG 60  
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40  
DB 61 CCATTGGCGATGGTGTCTCCATCGCGTACAGTCCATGACCAATACGCTACGACACGAC 120

AY033515  
 LOCUS AY033515 1119 bp DNA linear BCT 26-DEC-2001  
 DEFINITION Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (ispG) gene, complete cds.  
 ACCESSION AY033515  
 VERSION AY033515.1 GI:17978527  
 KEYWORDS  
 SOURCE Escherichia coli  
 ORGANISM Escherichia coli  
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.  
 REFERENCES  
 Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A., Arigoni, D., and Rohdich, F.  
 Studies on the nonmevalonate pathway to terpenes: The role of the Gcpe (ispG) protein  
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)  
 11752431  
 2 (bases 1 to 1119)  
 Rohdich, F., Hecht, S., Adam, P., Bacher, A. and Eisenreich, W.  
 Direct Submision  
 Submitted (27-APR-2001) Institut fuer Organische Chemie und Biochemie, Lehrstuhl III, Technische Universitaet Muenchen, Lichenberg Strasse 4, Garching D-85747, Germany  
 FEATURES  
 source  
 location/Qualifiers  
 1..1119  
 /organism="Escherichia coli"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:562"  
 1..1119  
 /gene="ispG"  
 1..1119  
 /gene="ispG"  
 /function="converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate, last but one step of the mevalonate-independent pathway for isoprenoid biosynthesis"  
 /note="ispG"  
 /codon\_start=1  
 /transl\_table=11  
 /product="1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase"  
 /protein\_id="AAK53460.1"  
 /db\_xref="GI:17978528"  
 /translation="MHQAPLORRKSRIYGVNPIDGAPIAVQSMTRTITDVEATVNOIKALERVADIVRSVPTMDAAFAFKLIKQOVNPLVADHFDRIALKYAEYVDCLIRINGINISNERIRMVVDCARDKNIPIRIGNAGSLEKIDOEKGEPTPOLLESAMRHVDLDRIPQPKVSKASDPLAVESRYLLAQIDOPHLGITTEAGARSQVKSATGLLSEGIIGDTRVSLADPVEIKVGFILKSLIRIRGINFTACPTCSKQEFDIVGTVALLEQLEDIITPMDSIIIGCVNNGEALVSTLGVTKGSKSGLYEDGV RKRLDNNMDIQLEARIKASQLEARRIDVQOVER"  
 BASE COUNT 272 a 280 c 315 g 252 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 9.1e-136 Length: 1119  
 Score: 1866.00 Matches: 372  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-921-992-50 (1-372) x AY033515 (1-1119)  
 QY 1 MethisanginalaProileGlnArgArgSerThrArgIleYrValGlyAsnVal 20  
 DB 1 ATGCATACCGGCTCCCAATTCACGTGAAATCAACACGATATTACGTTGGAAATGTG 60  
 QY 21 ProileGlyAspGlyAlaProileAlaValGlnSerMetThrAsnThrArgThrAsp 40  
 DB 61 CCGATTGGCGATGGTCCCATCGCGCTACAGTCATGACCAATACGCGTACACAGAT 120

QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyValAspIleValArg 60  
 DB 121 GTGAAGCAACGGTCAATCAATCAAGCGCTGGAAAGCGTTGGCGCTGATATCGTCCT 180  
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80  
 DB 181 GTCCGTCACCGACGATGAGCGCGGACGAGAACGCTTCAAACTCATCAAAACGAGGTTAAC 240  
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
 DB 241 GTCCCGCTGGTGGCTACATCCACTTCGACTACGATTCGCTGAAAGTAGCGGAATAC 300  
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120  
 DB 301 GCGCTGATGCTGTGCTATTAACCTGGCAATATCGTAATGAGAGCGATTCGATG 360  
 QY 121 ValValAspCysAlaArgAspLysAsnIleProileArgIleGlyValAsnAlaGlySer 140  
 DB 361 GTGGTTGACGTGCGCGGCGATAAACATTCGATCGATTGGCGTTAACCGCGATCG 420  
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
 DB 421 CTGGAATAAGATCTGCAGAAAGATATGGCAACCGACCGCGAGCGTTGCTGGAATCC 480  
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
 DB 481 GCCATCGCCATGTTATATCTCGATCGCTGAACCTCGATGATTCGATCAAGTCAGGTTG 540  
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
 DB 541 AAGCGCTGACCTCTTCCTCGCTGTGAGTCTTATCGTTTGTGCGCAAAACAAATCGAT 600  
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyValAlaLysSer 220  
 DB 601 CACCGCTGATCTGGGAGTACACGAAAGCGGAGCGCGCGACCGCGAGTAAATCC 660  
 QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
 DB 661 GCCATTGTTAGTCTGCTGCTGCTGTAAGGATGGACACACGCTGCCGATCACTG 720  
 QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
 DB 721 CGCGCGCATCGGCTCGAAGATCAAGATCGTTGATTTTGAATCGCTTCGATC 780  
 QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
 DB 781 CGTTCCGAGAGGTATCAACTTCATCGCTCGCACTGTTCCGCTCAGGAATTTGATGTT 840  
 QY 281 IleGlyThrValAsnAlaLeuGluGluArgLeuGluAspIleIleThrProMetAspVal 300  
 DB 841 ATCGGATCACTTAACCGCGCTGGACCAACGCTGGAAGATATCACTCCGATGAGACGT 900  
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGlyAlaLeuValSerThrLeuGlyVal 320  
 DB 901 TCGATTATCGGCTCGCTGCTGATGCGCGGAGCGCGCTGTTCTTCACTCGGCTC 960  
 QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValAlaArgLysAspArgLeuAsp 340  
 DB 961 ACCGCGCGCAACAAAGAACCGCTCTTATGAAGATGCGCTGCGCAAGACCGCTTGAC 1020  
 QY 341 AsnAspAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360  
 DB 1021 AACACGATATGATCGACCGCTGAGGACGCGATTCGTCGAAAGCCATCGACTCGAC 1080  
 QY 361 GluAlaArgArgIleAspValGlnGlnValGlyLys 372  
 DB 1081 GAAGCGGTGGAATTCGCTTACGACGAGGTTGAAAAA 1116  
 RESULT 2  
 LOCUS AX036302 1119 bp DNA linear PAT 16-NOV-2000  
 DEFINITION Sequence 29 from Patent EP1043403.  
 ACCESSION AX036302

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: November 23, 2003, 17:17:44 ; Search time 4608 Seconds

(without alignments)  
3302.600 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAPIQRKRSRIYGVN.....RAKASQDEARIDVQGVK 372

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n model -DBY=xlh  
-Q/cgn2.1/USFTO.spool/US0921992/rnatc\_21112003\_162546\_18406/app\_query.fasta.1.519  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pic -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US0921992 @CGN 1.1 3508 @runatc\_21112003\_162546\_18406 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEOUTRY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -MAIN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :  
1: gb\_ba :  
2: gb\_hcg :  
3: gb\_in :  
4: gb\_om :  
5: gb\_ov :  
6: gb\_pat :  
7: gb\_ph :  
8: gb\_pl :  
9: gb\_pr :  
10: gb\_ro :  
11: gb\_sts :  
12: gb\_sy :  
13: gb\_un :  
14: gb\_vi :  
15: em\_ba :  
16: em\_fun :  
17: em\_hum :  
18: em\_in :  
19: em\_mu :  
20: em\_om :  
21: em\_or :  
22: em\_ov :  
23: em\_pat :  
24: em\_ph :  
25: em\_pl :  
26: em\_ro :  
27: em\_sts :  
28: em\_un :

29: em\_vi :  
30: em\_hcg\_hum :  
31: em\_hcg\_in :  
32: em\_hcg\_other :  
33: em\_hcg\_mus :  
34: em\_hcg\_pln :  
35: em\_hcg\_rtd :  
36: em\_hcg\_mam :  
37: em\_hcg\_vtr :  
38: em\_sy :  
39: em\_hcg\_hum :  
40: em\_hcg\_mus :  
41: em\_hcg\_other :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1866	100.0	1119	1 AY033515	AY033515 Escherich
2	1866	100.0	1119	6 AX036302	AX036302 Sequence
3	1866	100.0	1119	6 AX038912	AX038912 Sequence
4	1866	100.0	1119	6 AX050487	AX050487 Sequence
5	1866	100.0	1119	6 AX393943	AX393943 Sequence
6	1866	100.0	1637	1 EC63PE	X64421 E.coli GGPB
7	1866	100.0	11521	1 AE005481	AE005481 Escherich
8	1866	100.0	13176	1 AE000338	AE000338 Escherich
9	1866	100.0	296827	1 AP002561	AP002561 Escherich
10	1866	100.0	300099	1 AE016764	AE016764 Escherich
11	1855	99.4	10225	1 AE015271	AE015271 Shigella
12	1855	99.4	290380	1 AE016987	AE016987 Shigella
13	1840	98.6	23647	1 AE008814	AE008814 Salmonella
14	1831	98.1	145050	1 AL627275	AL627275 Salmonella
15	1831	98.1	300431	1 AE016835	AE016835 Salmonella
16	1691	90.6	1010	1 AR027993	AR027993 Sequence
17	1691	90.6	1010	6 AR208328	AR208328 Sequence
18	1691	90.6	1010	6 BD084908	BD084908 Methods f
19	1645.5	88.2	301235	1 AP005075	AP005075 Vibrio pa
20	1631.5	87.4	301442	1 AE016798	AE016798 Vibrio vu
21	1624.5	87.1	1345	1 PSU67933	PSU67933 Providencia
22	1624.5	87.1	1345	6 AR027991	AR027991 Sequence
23	1624.5	87.1	1345	6 AR208326	AR208326 Sequence
24	1624.5	87.1	1345	6 BD084906	BD084906 Methods f
25	1622.5	87.0	11161	1 AE013738	AE013738 Yersinia
26	1622.5	87.0	204050	1 AJ414154	AJ414154 Yersinia
27	1621.5	86.9	11824	1 AE004161	AE004161 Vibrio ch
28	1603.5	85.9	10507	1 U32721	U32721 Haemophilus
29	1603.5	85.9	110000	6 AR274513_03	AR274513_03 Continuation (4 of
30	1593.5	85.4	1079	6 AE006237	AE006237 Pasteurel
31	1593.5	85.4	1079	6 AR027992	AR027992 Sequence
32	1593.5	85.4	1079	6 AR208327	AR208327 Sequence
33	1593.5	85.4	1079	6 BD084907	BD084907 Methods f
34	1584.5	84.9	10607	1 AE015769	AE015769 Shewanell
35	1372.5	73.6	17148	1 AE004798	AE004798 Pseudomon
36	1353	72.5	886	6 AR027995	AR027995 Sequence
37	1353	72.5	886	6 AR208330	AR208330 Sequence
38	1353	72.5	886	6 BD084910	BD084910 Methods f
39	1341.5	71.9	300861	1 AE016777	AE016777 Pseudomon
40	1335.5	71.6	311249	1 AE016860	AE016860 Pseudomon
41	1285.5	68.9	9608	1 AE014104	AE014104 Buchnera
42	1282	68.7	110000	6 BD061520_3	BD061520_3 Continuation (4 of
43	1282	68.7	347550	1 AP001118	AP001118 Buchnera
44	1226.5	65.7	349174	1 AB063522	AB063522 Wiggleleswo
45	1186	63.6	4775	1 D90881	D90881 E.coli geno

## ALIGNMENTS

RESULT 1

; SEQ ID NO 8947  
; LENGTH: 365  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-8947

Query Match 5.3%; Score 98.5; DB 15; Length 365;  
Best Local Similarity 20.9%; Pred. No. 0.24;  
Matches 77; Conservative 54; Mismatches 138; Indels 99; Gaps 19;

QY	18	GNVPIGGAPIAVQSMNTRTDVEA-----TVNQIKALERVGAD-----	57
Db	28	GSAPVGGGAVLHGTGEGAGVSVFARPGAYSTLRDLVDRFRAELVGADPFQREWLHR	87
QY	58	-----IVRSVPTMDAAE-----AFKLIXO-QVNVPLVADHFDYRIA--	94
Db	88	VWELDRHHEPLPTLGLIDIALMDLAGRYGEPVWRLLGGFRRAIPAYASTFSSVAEF	147
QY	95	LKVAEYGVDCLRINPGNIGNEERIRNVVDCARDKNIPIRIGVNGSLKDLQEKY----	150
Db	148	LDVAD--QCLALGYRGI---KLHAWGDARDDELCLALRDHVG---PDVPLMYDGSAG	197
QY	151	EPTPQAL-LESAMRHVDLRLNFDQKVSVDVFLAVESYRLAKQIDOPHLGITE	209
Db	198	FDLPDAIRLGRALSEADYL-----WYEEPIRE---FSISAYQRLAEAVDVPLVAETS	247
QY	210	AGG-----ARSGAVKSAIGLGLLSEGIQDTRLRVSLADPVEEIKVGFILKSLRIRS	262
Db	248	DGAHMAADPIRAGATFGVRAGTTLRGILTGAMRTAHLADAFR-----LRAEV	296
QY	263	RGI--NFIACTPSRQEF--DVIGTVNALQRLIEDIITPMDSIIGCVVNGPGEALVST	317
Db	297	HGSDIPVHHLQMAISNTTYESLVTSVNVVRERHVD-----DQGLVHAPA-GPGIALPLD	350
QY	318	LGVTGQNK 325	
Db	351	FGY--GNE 356	

Search completed: November 23, 2003, 17:22:06  
Job time : 50 secs

;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 5635  
;; LENGTH: 2368  
;; TYPE: PRF  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-5635

Query Match 5.3%; Score 99; DB 9; Length 2368;  
Best Local Similarity 20.8%; Pred. No. 4.1;  
Matches 89; Conservative 58; Mismatches 155; Indels 126; Gaps 19;

QY 3 NOAPIQ-RRKSTRIYGVNPIGIG-----APIAQSMNTNRTTVE---42  
DB 1532 NQAKTNIDSSIDBYVDNA-VKEGKAKINAVKTFSEYKDALAKIEAAVNAKVTEDNSN 1590  
QY 43 -ATVNOIKALERVGADIVRSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYG 101  
DB 1591 ASTSSIEAEKQKLAELKQTADQNVQATS---KODIEVQIHNDLD-----1633  
QY 102 VDCLRNPGNIGNEERIRNVDC--ARDKNIPIRIGVNAAGSLEKDLQEKYGEPTQALL 158  
DB 1634 ---NINDYTIPTGKESATTDLYAVADQKNNISADTNATQDEKQOAIKQVDQNVQTAL 1689  
QY 159 ESAMRHVDH-----LDRLNFD-----QFKSVKASDVFLAV 189  
DB 1690 ESTNNGVDNDVDALTOGKAALDAIQVDATYKPKANQVIDAKAEETKESIDSDQLTAE 1749  
QY 190 ESYRLA--KOIDPRLHIGTEAGARGAVKSAIGLLSEGIG-DTLRVSADPVE 246  
DB 1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKAQGLEAFNDIQIDSTEKQKA---1E 1804  
QY 247 EIKVGFDLKSLRISRGINFACPTCSROEPDVIGTVNALQORLEDTTPMDVSIIGCV 306  
DB 1805 ELETALDQIEA-----GVNVDADATTEKE-----AFTNALIEDILSKATEDISDQT 1850  
QY 307 VNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNNMDIDQLEA-RIR--AKASQDEAR 363  
DB 1851 TN---AEIATV-----KNSALEQLKAKRINPVVKKNALEAIR 1884  
QY 364 RIDVOQVE 371  
DB 1885 EVVNNQIE 1892

RESULT 14  
US-09-815-242-12389  
; Sequence 12389, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haeelbeck, Robert  
; APPLICANT: Ohlsen, Karl L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23  
;; PRIOR APPLICATION NUMBER: 60/253,625  
;; PRIOR FILING DATE: 2000-11-27  
;; PRIOR APPLICATION NUMBER: 60/257,931  
;; PRIOR FILING DATE: 2000-12-22  
;; PRIOR APPLICATION NUMBER: 60/269,308  
;; PRIOR FILING DATE: 2001-02-16  
;; NUMBER OF SEQ ID NOS: 14110  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 12389  
;; LENGTH: 2368  
;; TYPE: PRF  
;; ORGANISM: Staphylococcus aureus  
US-09-815-242-12389

Query Match 5.3%; Score 99; DB 9; Length 2368;  
Best Local Similarity 20.8%; Pred. No. 4.1;  
Matches 89; Conservative 58; Mismatches 155; Indels 126; Gaps 19;

QY 3 NOAPIQ-RRKSTRIYGVNPIGIG-----APIAQSMNTNRTTVE---42  
DB 1532 NQAKTNIDSSIDBYVDNA-VKEGKAKINAVKTFSEYKDALAKIEAAVNAKVTEDNSN 1590  
QY 43 -ATVNOIKALERVGADIVRSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYG 101  
DB 1591 ASTSSIEAEKQKLAELKQTADQNVQATS---KODIEVQIHNDLD-----1633  
QY 102 VDCLRNPGNIGNEERIRNVDC--ARDKNIPIRIGVNAAGSLEKDLQEKYGEPTQALL 158  
DB 1634 ---NINDYTIPTGKESATTDLYAVADQKNNISADTNATQDEKQOAIKQVDQNVQTAL 1689  
QY 159 ESAMRHVDH-----LDRLNFD-----QFKSVKASDVFLAV 189  
DB 1690 ESTNNGVDNDVDALTOGKAALDAIQVDATYKPKANQVIDAKAEETKESIDSDQLTAE 1749  
QY 190 ESYRLA--KOIDPRLHIGTEAGARGAVKSAIGLLSEGIG-DTLRVSADPVE 246  
DB 1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKAQGLEAFNDIQIDSTEKQKA---1E 1804  
QY 247 EIKVGFDLKSLRISRGINFACPTCSROEPDVIGTVNALQORLEDTTPMDVSIIGCV 306  
DB 1805 ELETALDQIEA-----GVNVDADATTEKE-----AFTNALIEDILSKATEDISDQT 1850  
QY 307 VNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNNMDIDQLEA-RIR--AKASQDEAR 363  
DB 1851 TN---AEIATV-----KNSALEQLKAKRINPVVKKNALEAIR 1884  
QY 364 RIDVOQVE 371  
DB 1885 EVVNNQIE 1892

RESULT 15  
US-10-156-761-8947  
; Sequence 8947, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; PRIOR FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109

```

Qy      330 -----YEDGVKRDRLDNNMDIDQLEARIRAKASQLDEARRIDV 367
          | : : : : | : : |
Db      517 SVEADAGMQQEGAGISDQETQATEEVEKVEVSVEADAGMQQEL--VDV 561

```

```

RESULT 11
US-10-306-905-11
; Sequence 11, Application US/10306905
; Publication No. US20030167513a1
; GENERAL INFORMATION:
; APPLICANT: MOURAD, GEORGE S.
; APPLICANT: JUNK, DAMIAN J.
; TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)
; TITLE OF INVENTION: MUTANTS DESENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
; TITLE OF INVENTION: CONTROL
; FILE REFERENCE: 3220/93981
; CURRENT APPLICATION NUMBER: US/10/306,905
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/339,895
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 523
; TYPE: PRF
; ORGANISM: Escherichia coli
; US-10-306-905-11

```

```

: CURRENT FILING DATE: 2002-11-27
: PRIOR APPLICATION NUMBER: 60/339,895
: PRIOR FILING DATE: 2001-11-30
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 15
: LENGTH: 573
: TYPE: PR1
: ORGANISM: Streptomyces coelicolor
: US-10-306-905-15

```

```

RESULT 12
US-10-306-905-15
; Sequence 15, Application US/10306905
; Publication No. US20030167513A1
; GENERAL INFORMATION:
; APPLICANT: MOURAD, GEORGE S.
; TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)
; TITLE OF INVENTION: MUTANTS DESENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
; TITLE OF INVENTION: CONTROL
; FILE REFERENCE: 3220/93381
; CURRENT APPLICATION NUMBER: US/10/306,905

```

```

RESULT 13
US-09-815-242-5635
Sequence 5635, Application US/03815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931

```





US-10-306-905-16

Query Match	6.8%	Score 126;	DB 12;	Length 533;
Best Local Similarity	23.4%;	Pred. No. 0.00066;		
Matches 90;	Conservative 57;	Mismatches 115;	Indels 122;	Gaps 22

```

QY 14 RIYGVNVPJGDGAPILAVOSMTNTRTDDYATNOJKALEKRGAIIVRSVYPTMDAA--EA 71
Db 7 RIIIFPTLRDGE---QSPGATLT--VEEKLSIRALLARLGVDDIIKGGFPFSPGDFEA 60
QY 72 FKLIKOOV---NVPLV-----ADIHDPYRIATLK 96
Db 61 VQCIAGTGVGEMGNPVI CGLARATQDKIAKAALKPAKAKHRIHTEFLATSDIHLEHKIKT 120
QY 97 VAE-----YG---VDCIRINPNI G--NEERIMVYDCAKDKNIPRIGVNAGSL 141
Db 121 RAEVLAI VEPMAVYAKSLVNDIEFSPEDAGRSDFEFLQVLEAA-----ISAGAT 170
QY 142 EKLOEYKEGEPRTQ---ALLSASMRVHDLRLNFDQKSVYKA--SPVFLAVESRYLLAK 197
Db 171 TINIPIPVGTTPIPAEGLIKGIADNVP---NIDQAIISVHGHNDDGLAVANFLEAYK 225
QY 198 QIDOPHLGLITENGARSAGVAKSAIGLILSEGIGDTRLVSLADPAVEIKYVFDILKS 257
Db 226 -----NGARQ--LECTI-----NGIGE--RAGNNA--LEE-----LVMA 253
QY 258 LRIIRSGIN FLACPRCSROEPDVI GTVA--LEQLEDDITIPMDVSIIGCVNNGPEALV 315
Db 254 LHVRSYFNFELCRPADSTAPLTNIDITKIIYATRSVLSELTGMVVO-----PNKAI V 305
QY 316 STLGVTGNKKSGLYDEGVKRDRL 339
Db 306 ---GANPAHESGIHODGVLLKNTL 326

```

## RESULT 7

```

US-10-306-905-14
: Sequence 14, Application US/10306905
: Publication No. US20030167513A1
: GENERAL INFORMATION:
: APPLICANT: MOHRAD, GEORGE S.
: TITLE OF INVENTION: JUNK, DAVIAN J.
: TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)
: TITLE OF INVENTION: MUTANTS DESENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
: FILE REFERENCE: 3220/93981
: CURRENT APPLICATION NUMBER: US/10/306,905
: CURRENT FILING DATE: 2002-11-27
: PRIOR APPLICATION NUMBER: 60/339,895
: PRIOR FILING DATE: 2001-11-30
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 14
: LENGTH: 533
: TYPE: PRT
: ORGANISM: Microcystis aeruginosa
US-10-306-905-14

```

	Query Match	Similarity	5.9%	Score	110.5	DB 12	Length	533	
	Best Local	86	Conservative	Matched	No.	0.026	Mismatches	128	Indels 145; Gaps 21
Oy	14 RIYGVNVIGDAPAVIVOSMTNTRTTTDEATNQIKALERVGADIVRVSVP-----	64							
Dd	7 RVIIFDTTLRDOE-----QSFGAALNVDELKLT--ARALLRLGVDVIEAGFPIASPDFFEA	60							
Oy	65 -----TWDAAEAFK-LIKQOVNVL-VADIHPDYRI- --	93							
Dd	61 VQKIAGSVGEADSPILCGLARATQDKIRKSADDALRPAAKPRIHFLATSIDLGYKLKK	120							
Oy	94 -----ALKVAIYGVDCLRNPNIGTG--NEBRIMVVDCARDKNIPIRIGNMGS	140							
Dd	121 TRQEVLEIVEPMWAAVAKSFLLNDIVERSPEBAGSGSDFELVQLVERA-----IAAGA	170							

```

QY      14  LEKDLOEKGEPTPO---ALLESAMRHVDHLDRINFDOFKVSVA--SDVFLAV---ESY 192
Dd      171  TVNINIDTVGYTTPSEFGALIRGIENVP-----NIDQAIISVGHDDLGIAVANFLEAV 225

QY      193  RLIAKOIDPPLHLGITTEAGARGASGAVKSAIGLGLLSEBGIDTLRVSLADPVEEIKVGF 252
Dd      226  KNGAROLETETIN-GIGERAGNAS-----LEE-----250

QY      253  DILKSLRIRSRGIN-FIACPTCSROEPDVGTVNALE--QRLEDIITPMVDYSIIIGCVNCP 310
Dd      251  -LWMAHVRKSRFENPFLGRPAESTPLPKRINTKELYRISRLVSNULTGMIYQ-----P 301

QY      311  GEALVSTLGVTCGNKKSGLYEDGYVKORL-----DNNDMIDOLEARIRAKAS 357
Dd      302  NKRAIV---GANNFAHESGIHODGVKHLKLTYEIMDAESIGLTNNQIYVLGKLSGR--NAFRS 357

QY      358  QLDE 361
Dd      358  RLQE 361 --

```

## RESULT 8

```

US-10-156-761-10088
: Sequence 10088, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HIROO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156,761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001.204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-06-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 10088
: LENGTH: 1046
: TYPES: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-10088

```

Query Match	5.8%	Score 108	DB 15	Length 1046	
Best Local Similarly	21.8%	Pred. No. 0.13			
Matches 86	Conservative 48	Mismatches 139	Indels 122	Gaps 18	

```

QY      11 KSRIYVAVNPJEDGCA-PRAYOSMNTPTTDEAVTNOIKALERGADIVAVSPTMDAA 69
Db      614 KSDIDIALVVAANDGCMPTOTIEALNHAAQADVPIV-VA NKIDVDEGAD-----PT----- 662
QY      70 EAFELIKQOVNPELVADHFDYRIALKVAEYGVDCLRINPGNIGNEIRIMVDCARDKN 125
Db      663 ---KVGGOLTEFGLV-----EYEGD-----TWFPV----- 686
QY      130 IPIRIGVWNGSL-----EKDLOEKYGGPTQALLSESAMRHVDLDTNPDQKV- 178
Db      687 ISAKQGNIBESLEAVVLTADASLDRANPEODAGIATES-----HLDGRCAVAVTL 740
QY      179 ----SVKASVPLAVESYRLAKOIIDPRLHIGTEGARGASGAYSAIGLLLSBIGD 234
Db      741 VQRGTLRVGPVTWVGDAYGRVAMLDKCE-NVEAGSGSTPLV-----LGLTVNPGAGD 794
QY      235 TLKAVSLADPVEBIKQGPDLKSLKIRSGINFI-----ACPTCSQNEPDI 281
Db      795 NFLV-----VEDRIRARQIAKRAARENNAPFARGVRFSLENUDELKAGLQVQENLI 848

```

```

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5705
LENGTH: 378
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5705

```

Query Match	40.8%;	Score 761;	DB 10;	length 378;
Best Local Similarity	44.3%;	Pred. No. 2.7e-69;		
Matches 159;	Conservative 64;	Mismatches 124;	Indels 12;	Gaps 3

Qy	9	RRKSTRILYGVNVPJGSDGAPFLAVOSMTNTRFTTDEATVNOIKALERGAQIVRVSPTTMA	68
Db	5	RRKTRQLMVGKVGSGSDHPISVQSMTTTKKHNDINGTLNQIQAOLTLATGCDIVRACKXTYD	64
Qy	69	AEAPFKLKOQVNVPLVADHIFEDYRIALKVAEYVDCILRINPNIYN-EERIRNVDCARD	122
Db	65	AEALPPIIAKSSPIFVIADHIFQPKYTFPAALDACCAAVRVPNGIKSFQDGVKVAKKAAD	121
Qy	128	KNPIRIGVAGSLSEKLOEKY-GEPTPQALLSEAMRHVDHRLNFQDFKYSVKASDVCF	18
Db	125	AGPIRIGVAGGSLDKRILDKYHGKATPPEALVESAMMENGLPFEHFGQIALISVKHSIDV	18
Qy	187	LVNESHLLAKQIDQPHLGLTTEGARGSAVNSAIGLGLLSEEGIDTLRVSLLADPYE	24
Db	185	LMVEAYRQLAEQSQSYPHLGVTEAGRPFMOTIKSSVAFGLLSQGGDITIRVSLSDPYE	24
Qy	247	EIKVGFPIILSLRSRGINFIACTPSROEPIVIGTVNLAEOLEDITTPMDVSIIGCV	30
Db	245	EIKVGDDIIGLSINRPRKLEIVSCPSGGRQVVDVYLSAEVTEALDGMETPLRVAMGCV	30
Qy	307	VNGGELAVSTLGYTGANKKSGLYEDGVCRDRLDNDMDIDOLEARIRAKASQULDEARRI	365
Db	305	VNGGEGARDPDLGVAASNGKQGVFVKGVEVIRKIVPEQIYEVL-----IEEMRI	353

RESULT 5  
 US-09-712-256  
 Sequence 256, Application US/09712363  
 Patent No. US20020164588A1  
 GENERAL INFORMATION:  
 APPLICANT: Eisenberg, David  
 APPLICANT: Rotstein, Sergio H.  
 APPLICANT: Marcotte, Edward M.  
 TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND  
 TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS  
 FILE REFERENCE: 07419-012001  
 CURRENT APPLICATION NUMBER: US/09/712,363  
 CURRENT FILING DATE: 2000-11-13  
 PRIOR APPLICATION NUMBER: PCT/US00/02246  
 PRIOR FILING DATE: 2000-01-28  
 PRIOR APPLICATION NUMBER: 60/179,531  
 PRIOR FILING DATE: 2000-02-01  
 PRIOR APPLICATION NUMBER: 60/117,844  
 PRIOR FILING DATE: 1999-01-29  
 PRIOR APPLICATION NUMBER: 60/118,206,

```

? PRIOR FILING DATE: 1999-02-01
? PRIOR APPLICATION NUMBER: 60/126,593
? PRIOR FILING DATE: 1999-03-26
? PRIOR APPLICATION NUMBER: 60/134,093
? PRIOR FILING DATE: 1999-05-14
? PRIOR APPLICATION NUMBER: 60/134,092
? PRIOR FILING DATE: 1999-05-14
? PRIOR APPLICATION NUMBER: 60/165,124
? PRIOR FILING DATE: 1999-11-12
? PRIOR APPLICATION NUMBER: 60/165,086
? PRIOR FILING DATE: 1999-11-12
? NUMBER OF SEQ ID NOS: 292
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 256
? LENGTH: 387
? TYPE: prt
? ORGANISM: Mycobacterium tuberculosis
US-09-712-363-256

```

Query Match	40.3%	Score 751.5;	DB 10;	length 387;
Best Local Similarity	42.7%	Pred. No. 2.6e-68;		
Matches 156;	Conservative 64;	Mismatches 134;	Indels 11;	Gaps 2

```

Qy  RRKSTIYGVNPIDGGAPIAVOSWTRTDTDEATVNOIKALERVADIVRVSPTMDA 68
Dh  18 RRATROLMGNGVGSHPHVSVOSSCTTKTHDNVSTIOQLAELTAAACDILVRVACRQED 77
Qy  69 AEAFLIKOOVNPVLADIHEDYRIALKVAEYVDCLRINFGNIGN-EEBIRMVVDICARD 127
Dh  78 ADALAEIARHSQIPVADIHFEQPRYIFAIIDAGCAAVRVNPNKIEFEDGVRGEVAKAGA 137
Qy  128 KNIPRIPIGNAAISLEKDLQEKKGTEPTPALLSASMRHVBDLDRINFPQFVSVKASDVFL 187
Dh  138 AGPIRPIGNASLDRKFMEXKGKATPAPALVESALMEASLFEHGFSDIKI SVKHANDPVV 197
Qy  188 AVESYRLIAKQIDQDPLHIGITEAGARAGVAKSAIGLLISEGIGDTLRVLSAADPVEE 247
Dh  198 MVAAYVELLARCDDYPLHLGVTBAGAPFOGTIKSAVAGALLSRGIGTTRVLSAPVEE 257
Qy  248 IKVGFIDILKSLIRNSGINFIACPTCSRQEHFVIGTVNALEQRLEDIITPMVYSIIGCVV 307
Dh  258 VKVGNQVLESTNLPRPSLEIVSCBPCGQAVDYVTTLANBVTAGLDGIDVPLRVAVNGCVV 317
Qy  308 NGEGEALVSTLGVTGNGKKSGLYEDGVAKRDLNNNDIMIDOLEARIRAKASQDLEARDIV 367
Dh  318 NGEGEARADIGVASNGKGOIFVRGEVITKTPBAQIVETL-----IEAMRLAA 367
Qy  368 QQVEK 372
Dh  368 EMGEQ 372

```

```

RESULT 6
US-10-306-905-16
: Sequence 16, Application US/10306905
: Publication NO. US20030167513A1
: GENERAL INFORMATION:
: APPLICANT: MOURAD, GEORGE S.
: APPLICANT: JUNK, DAMIAN J.
: TITLE OF INVENTION: SELECTION AND USE OF ISOPRENYLAMATE SYNTHASE (IPMS)
: TITLE OF INVENTION: MUTANTS DESENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
: TITLE OF INVENTION: CONTROL
: FILE REFERENCE: 3220/93981
: CURRENT APPLICATION NUMBER: US/10/306,905
: CURRENT FILING DATE: 2002-11-27
: PRIOR APPLICATION NUMBER: 60/339,895
: PRIOR FILING DATE: 2001-11-30
: NUMBER OF SEQ ID NOS: 45
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 16
: LENGTH: 533
: TYPE: PR1
: ORGANISM: Synechococcus PCC6803

```

[illegible]

```

RESULT 2
US-09-881-752A-196
; Sequence 196, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; PRIOR FILING DATE: 1997-04-01
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 196
; LENGTH: 359
; TYPE: Prt
; ORGANISM: Helicobacter pylori
; US-09-881-752A-196

```

Query Match	42.6%	Score 794	DB 10	Length 359
Best Local Similarity	45.8%	Pred. No. 1e-72		
Matches 162	Conservative	73	Mismatches 117	Indels 2
				Gaps 2

QY 9 RRKSRIRVGNPIDDGA PIAVOSMTNTRFTTVEATVNO IKLBERGADIVAVSVTMDA 68  
 Db 5 RVRTQJIRIGVIAIGDAP ISTQSMTFSTADIESTRKQIDRLKAGADLVAVASNEED 64  
 QY 69 AEAFLIKQOVNVPVADIHFDYRIALKAEXGVDCLRINPGNIGNEERIIRMVDCARDK 128  
 Db 65 ALALKEIKKVPPLPLIADIHFFHYKPAL -IAOSVDAIRINPGNISGEKERIKKAVDDCKEK 123  
 QY 129 NIPIRIGNAGSLEEDLOEKYGEPTPOALLGESAMRHVHDLBTLNPOGFVSVKASDVFLA 188  
 Db 124 NIPIRIGNAGSLEKQFDOKIG-PTPKGVESALYNAKLLEDDIDFNFKISLKASVITYET 182  
 QY 189 VESYRLAKQIDQPLHLGITTEAGARGSGAVKSAIGLULLSEGIQDITLVLVSLAADVEEII 248  
 Db 183 IEAYMLPLVYYPHLGVTEAGNLFSSSIKAMALGILMIGICDITMVSITGELENEI 242  
 QY 249 KVGFDILSLRIRSGINFIACPTCSROBFVIGYTNALOEQLEDIITMDVSIIGCVN 308  
 Db 243 KVARAILHSGRLKKGKGINISCPITGRLEANLVDAIKVEKLSHITKTLDDISVMGCVVN 302  
 QY 309 GPGEALVSTLGVTVGNKKSGLYEDGVDRKDLNNNDIMIDGLEARIKASQOLDEA 362

```

Db      303  ALGEAKHADMAIAFGNRSGLIKEGVIIHKLAEKDLFETFVIEVNLAKEREKS 356
          ||| : : || : : : | : : | : : :

```

```

RESULT 3
US-10-156-761-9186
: Sequence 9186, Application US/10156761
: Publication No. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMODA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHITUKU
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156, 761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 9186
: LENGTH: 385
: TYPE: PRT
: ORGANISM: Streptomyces avermitilis
US-10-156-761-9186

```

Query Match	42.1%;	Score 786.5;	DB 15;	Length 385;
Best Local Similarity	44.4%;	Pred. No. 6.6e-72;		
Matches 164;	Conservative 71;	Mismatches 121;	Indels 13;	Gaps 4;

```

QY 6 PI-ORRSTRUYGNVBIIGCAPIAVOSMTTPTTDEATVNOIKALERVADIVRVSVP
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 15 PIARRRSRQIQVPEPLVVGTAIPVSVOSMTTTRTISDQIGATIQOIAELTASGQIVRACP 74
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 65 TMDAAEFKLIKQOVNPLVADIHFDRIALKVAEVGDCLIRINPGIGN-EERI RVVD 123
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 75 TQDDADALAVIARRSQIPVADIHFGQRYVPAALIEAGCAARVNPNGNIKQDDDKVKEIAR 134
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 124 CARDKNTPIRIGNVAGSLEKDLOEKYEPPTPOALLBESAMRHVDHLDRINPOQFVSVKAS 183
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 135 AAKHGHPRIPIGVAGSLDRILLEKYKATPEALVESALMEASLFEEDHFDIKISVGN 194
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 184 DVFLAVESYLLAQQIDQPLHIGITTEAGARSAGVKAIGLISGTLRPLVLAAD 243
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 195 DPVVAVNAVAYQOLAQQCPPLHLGTVEAGPAQGTIKSAVARFALLBSGIGTIRVSJAP 254
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 244 PVEIRIKGFOLIKSLIRSRGINFIACPTCSROEFYDVGYNVNALEQRLEDIITTPMDYSII 303
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 255 PVEIRIKGIGQLBESLGRQRRLLEIVSCPSCCRADVYKLAEBYTAGLEGNEVPLRVAVM 314
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 304 GCVVNGEGEALVSTLVTGANGKSGLYEDGVKRDLDNNNDMI DQLEAIRAKASQOLDEAR 363
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 315 GCVVNGEGEARADLIGVASGNGKQOIFVKGSEVITKTVESKIVETL-----IEEAM 364
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 364 RIDVQVKEV 372
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 365 KI-AEQMEQ 372
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

US-09-738-626-5705  
 : Sequence 5705, Application US/09738626  
 : Publication No. US2002019765A1  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: NAKAGAWA, SATOSHI  
 : APPLICANT: MIZOGUCHI, HIROSHI  
 : APPLICANT: ANDO, SEIKO  
 : APPLICANT: HAYASHI, MIKIRO

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:16:44 ; Search time 48 Seconds

(without alignments)  
1414.836 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866  
Sequence: 1 MHQAQPIQRKSTRIVGVN.....RAKASQDEARRIDVQVEK 372

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications\_AA:\*  
2: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	808	43.3	385	US-10-156-761-10098	Sequence 10098, A
2	794	42.6	359	US-09-881-752A-196	Sequence 196, App
3	786.5	42.1	385	US-10-156-761-9186	Sequence 9186, Ap
4	761	40.8	378	US-09-738-626-5705	Sequence 5705, Ap
5	751.5	40.3	397	US-09-712-363-256	Sequence 256, App
6	126	6.8	333	US-10-306-905-16	Sequence 16, App
7	110.5	5.9	533	US-10-306-905-14	Sequence 14, Appl
8	108	5.8	1046	US-10-156-761-10088	Sequence 10088, A
9	105.5	5.7	578	US-09-159-469-50	Sequence 50, Appl
10	105.5	5.7	578	US-09-798-042-50	Sequence 50, Appl
11	101	5.4	523	US-10-306-905-11	Sequence 11, Appl
12	99.5	5.3	573	US-10-306-905-15	Sequence 15, Appl
13	99	5.3	2368	US-09-815-242-5635	Sequence 5635, Ap
14	99	5.3	2368	US-09-815-242-12389	Sequence 12389, A
15	98.5	5.3	365	US-10-156-761-8947	Sequence 8947, Ap

16	96	5.1	504	9	US-09-815-242-10969	Sequence 10969, A
17	96	5.1	504	15	US-10-260-877-70	Sequence 70, Appl
18	94.5	5.1	290	9	US-09-815-242-5590	Sequence 5590, Ap
19	94.5	5.1	295	9	US-09-815-242-12428	Sequence 12428, A
20	94.5	5.1	295	9	US-09-815-242-12428	Sequence 12428, A
21	93	5.0	483	12	US-10-301-997-51	Sequence 51, Appl
22	93	5.0	459	9	US-09-815-242-12703	Sequence 12703, A
23	93	5.0	1004	10	US-09-738-626-5676	Sequence 5676, Ap
24	93	5.0	2478	9	US-09-815-242-5816	Sequence 5816, Ap
25	93	5.0	2478	9	US-09-815-242-12967	Sequence 12967, A
26	92.5	5.0	307	15	US-10-128-714-4018	Sequence 4018, Ap
27	92	4.9	291	9	US-09-815-242-11266	Sequence 11266, A
28	92	4.9	498	10	US-09-738-626-5600	Sequence 5600, Ap
29	92	4.9	1798	9	US-09-845-583-8	Sequence 8, Appl
30	91.5	4.9	297	9	US-09-815-242-13258	Sequence 13258, A
31	91.5	4.9	332	9	US-09-815-242-13072	Sequence 13072, A
32	91.5	4.9	342	12	US-10-166-225A-63	Sequence 63, Appl
33	91.5	4.9	589	12	US-10-306-905-5	Sequence 5, Appl
34	91.5	4.9	756	10	US-09-738-626-6091	Sequence 6091, Ap
35	91	4.9	652	15	US-10-102-806-667	Sequence 667, App
36	91	4.9	855	15	US-10-156-761-14612	Sequence 14612, A
37	91	4.9	1798	10	US-09-938-275-9	Sequence 9, Appl
38	90.5	4.8	455	9	US-09-815-242-5420	Sequence 5420, Ap
39	90.5	4.8	455	9	US-09-815-242-11218	Sequence 11218, A
40	90.5	4.8	477	9	US-09-815-242-11826	Sequence 11826, A
41	90.5	4.8	502	15	US-10-156-761-12534	Sequence 12534, A
42	90.5	4.8	584	15	US-10-156-761-13475	Sequence 13475, A
43	90.5	4.8	842	9	US-09-815-242-11950	Sequence 11950, A
44	90	4.8	233	10	US-09-971-536-71	Sequence 71, Appl
45	90	4.8	358	9	US-09-784-508-4	Sequence 4, Appl

#### ALIGNMENTS

RESULT 1  
US-10-156-761-10098 Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 10098  
; LENGTH: 385  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-10098  
Query Match 43.3%; Score 808; DB 15; Length 385;  
Best Local Similarity 44.9%; Pred. No. 4,1e-74;  
Matches 164; Conservative 70; Mismatches 119; Indels 12; Gaps 3;

QY 8 ORKRSRIYGVNPIGDGAPIVQSWTNRTRTVDVEATVQIKALEVGVADIVSVPTMD 67  
DB 17 ERKRSQIOGVYAVGADAPVQSVQSTTTTRISDIGATLQOIMELTISGCOIVAVCPDQ 76  
QY 68 AAEAFPLIKQOVNPLVADIDHFDYRIALKVAEYGVDCIRINPGNIGN-ERRIMVVDCAK 126  
DB 77 DADALAVIRAKSQIPYIADIHFDQPKYVFAALTEGCAAVVNGNINQPDFKYEIKAAK 136

```
; Sequence 2, Application US/09612964
; Patent No. 6403342
; GENERAL INFORMATION:
; APPLICANT: Gusevskiy Mikhail Markovich
; APPLICANT: Lunts Maria Grigorievna
; APPLICANT: Kozlov Yuriy Ivanovich
; APPLICANT: Ivanovskaya Lirina Valerievna
; APPLICANT: Voroshilova Elvira Borisovna
; TITLE OF INVENTION: DNA CODING FOR MUTANT ISOPROPYLMALATE SYNTHASE
; TITLE OF INVENTION: L-LEUCINE-PRODUCING MICROORGANISM AND METHOD FOR PRODUCING
; TITLE OF INVENTION: L-LEUCINE
; FILE REFERENCE: 193845050
; CURRENT APPLICATION NUMBER: US/09/612,964
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: RU 99114325
; PRIOR FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-612-964-2
```

```
Query Match 5.4%; Score 101; DB 4; Length 523;
Best Local Similarity 20.3%; Pred. No. 0.046;
Matches 83; Conservative 62; Mismatches 142; Indels 122; Gaps 18;
```

```
QY 12 STRIYGVNVPICGAPFAVDSMTTRTTDEATVNOIKALERVAGADIVRSVPTMDAA-- 69
Db 2 SQOVIIIFDTLRDGEQALQASLS-----VKEKQLALALERMGVDMVEVGFVSPGDF 55
QY 70 EAFELIKQOVN---VPLVADIHFDYRIA---LKVAEYGVDCURI-----NPGNIGNE- 115
Db 56 ESVOTIRQVKNRSVCLACVEKIDVALESKVAE---AFRIHTFATSPMHIATKL 111
QY 116 -----ERIRNVVDCAR-----DKNIPRIGVNAAGSLEKDLQEK 148
Db 112 RSTLDEVIERAIVMKRARNYTDVDFSCBDAGRTPIADLARVVEAIVAGATTINIPDT 171
QY 149 YGEPTP---QALLESAMRHYDHLDFQPKYSVKA-SVFLAVESYRLAKOIDOPLH 204
Db 172 VGYTMPPEFAGIISGLYERVP---NIDKAIISVHTHDDLGLAVGN-SLAA-----VH 219
QY 205 LGITEAGGARSAGVKAISAGLLSEGIQDPLRVSLADPVEIKVGFDTLKSLRIRSRG 264
Db 220 AGARQVEGANN-----GIGERAGNCSLEEVIMAIKVRKDI----- 255
QY 265 INFIACTCSROEDVIGTVNALE-ORLEDIITPMVSIIGCVVNGPEALVSTLGVTTG 323
Db 256 -----NVHTAINHQEIWRTSQVSO-----ICNMPIPANKKAIVSGAF 293
QY 324 NKSGLYEDGVKRDRLDNDNM-----IDOLEAIRAKASOLDEARRID 366
Db 294 AHSSGIHQDGLVKNRENYEIMTPESIGLNQIOLNLTSGRAAVKHRMD 342
```

```
Search completed: November 23, 2003, 17:17:39
Job time : 23 secs
```

```

QY      277 EFDVIGTVAL-----EQLEEDITPMADVSIICGV--VNPGELATSTLGTWGNKKSGL 329
Db      457 ELFGVEAAEAIVPSLEENKTGEEVVAPAEAOQLSAPSVSAIPVGSESTVLGTBTGDLSKEV 516
QY      330 -----YEDGVRRKDRLDNNMDIOLLEARIRAKASQDLDEARRIDV 367
Db      517 SVEADAGMOEGEAGISDOETQATEEVKEVESVEDADMGOEL--VDV 561

RESULT 13
US-09-106-582-50
; Sequence 50, Application US/09106582
; Patent No. 6306402
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPlicant: Iodes, Michael J.
ApplicANt: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMpuTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/106,582
FILING DATE: 29-JUN-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE / DOCKET NUMBER: 210121.439c2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAx: 206-682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 578 amino acids
TYPE: amino acid
STRANDEDNESS: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-106-582-50

Query Match          5.7%; Score 105.5; DB 4; Length 578;
Best Local Similarity 19.5%; Pred. No. 0.018;
Matches    91; Conservative   77; Mismatches 176; Indels 123; Gaps 21;

       7 IQRRKSTRIFYGNVP-----IDGA-----PIAVQSMTNTRT-TDEVATYNQ- 47
           :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     112 VQEESTGYMYLINAFKAVRFFRIEKSKAHEEPQTVDPSVESASTGSVDTPGEODEIQE 171
               |::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      48 IKALERVGD-----IVRSVPTMDAAEF-KLIQQQNVP 82
           :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     172 APAIEVEETEBOEVIIEBGLTLIDLPQPVAQPVPAEALPGEVEAEAIIVPSLEENKOEV 231
               ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      83 LVADIHPDRILTKV---AEYGVDCLRINGNIENGERRIMRVNDVCARDKNIPRIIGNVAG 139
           ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     232 VAPEAQOLESAPESAPAPQESTVLGVAEGDKSEVSVEANDVAQRK-----VISG 284
               ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY     140 SLKKLOEKY-GERTPOLLE-SAMRHVDHLRLNFDFKVSY----- 180
           ::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     285 QOEGERIAELEGGEARAVEVEETREVLLKBDLTLDLEQPVNAQPPVAEALPGEVEAEAIV 344
               ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY     181 ----KSDVFIAVESYRL-LAKOIDPLH-----LGITEBAGARGCAVKSAIGL---- 224

```

```

Db      345 PSLEENKLGQEVVVAPEAQQLSAPVSPAPQPESTYUGVTE-----GDKSEVSEADA 398
Oy      225 GLLISEGIGDTRLRVSLAADPVEEIKVGFED-----ILKSIRSRGINFIACPTCSRQ 276
Db      399 GMOOEAGISD--DETQATEEVEKEVSEVETKTEPEVILEEGTLIDLECPVAQVPVAAEA 456
Oy      277 EFDVIGVYVNL-----EQRLEDITPMDVSIIGCV--VNGPGALVSTGVTGNKSKGL 329
Db      457 ELPGVEAAEAIVPSLEENKLGQEVVVAPEAQQLSAPVSPAPQPESTYUGVTEGDLKSEV 516
Oy      330 -----YEDGVKRDRLDNNMDIMDLKARIRAKASQDLKARRIDV 367
Db      517 SVENDAGMOOEAGISDDETQATEEVEKEVSEVSEADAGMOQL--VDV 561

RESULT 14
; Sequence 29387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ ID NO 29387
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29387

Query Match      5.4%; Score 101, DB 4; Length 465;
Best Local Similarity 21.8%; Pred. No. 0.038;
Matches 81; Conservative 43; Mismatches 122; Indels 126; Gaps 15;

Oy      25 GAPAAV-QSMTNRTTDDVEATVNOIKALERVGADIVRSVPTMDABAPFLIKQOVNVP 83
Db      59 GLEIALLETTVALALDLDLE---EDRDNRKVGEDL-----QGDAAVARGAVDQOVQPAQ 108
Oy      84 VADIHFDYRIALKAVEYGVDCLRINPGNISNEERIRMVNDCA----- 125
Db      109 LLE-----RLAMP-GDAGVDLVVV--GVRRHHLHAAVNVQALHAGSIDVUGGGDVLDAFA 160
Oy      126 -----RDKNIPTIR---IGVNAQSLEKIDQOEKKGTEPTALLBSA- 161
Db      161 MVEAFDEFLDRLGIVGFVFDRLDADLPARCGHGAEOAGELAFDVEVNLAEVSGALVETGP 220
Oy      162 -----MRHVDHLDRLL-----NPGQFVSVKASDVFLEAVESYRLARQIDQRLH 204
Db      221 DVHLAALDVGVQVAVDVGQADRVAVVGAAPFELERVDVTDAAVAAVADETIQRTADAPQACD 280
Oy      205 LGITEAGGA--RSQAVKSAIGLGL--LLSEGIQDTRLRVSLAADPVEEIKVGFEDILKSLR 259
Db      281 VQLAEVAVNAHQIGTLGFGVGGGLSVLHREGHGAGARVALBEE----- 324
Oy      260 IRSRGINFIACTPSCROEPFVIGTVALLEQRLEDITPMDVSIIGCVVNGPGSA----- 313
Db      325 -----FVDMPGRAAVEHDV-----DIVLLEQEDFLGTMLGGFGEAHVHEQL 365
Oy      314 --LVSTLGTGTGG 323
Db      366 AOLLDALGVGGG 377

RESULT 15
US-09-612-964-2

```

[illegible]

QY 134 ICNV-----AGSLKDLQEKYGEPTPOLLESANRHHVLDRLNFDQFKVSVKASDVFLA 188  
DB 414 --INHNALSGTTRIPLSDAEOFPFGRSLVE-----ILLRTPPVFLV 453  
QY 189 VESYRLAKQIDQPLHIGITEAGARGSAIKSAGLGLLSEGD----- 234  
DB 454 DEIDGLKSAIDS-----SVLE---LRIGVGNDSPLRLFNHVGKSHIPGIDRSSLYRV 506  
QY 235 -----TLRVSLAAD-----PVEIKVGFDLKSLIRSRGINFIACPTCSROEF 278  
DB 507 NMAIMVCTLSVSLQMLNQCERFEMPIKE--TWSDADDTITQALGI-----TRETLEY 559  
QY 279 DVIGTVNALEQRLEDITTPMDVSIIGCVN--GPGELVSTLGVGKNGKSGLYEDGVR 335  
DB 560 LFSNDJGQLPFHILQ-----SLSRNLANWRSNDGHRLETTLATLGNLIADLSSGRL 611  
QY 336 KQRLDNNMDIDQLEARIKAKAS 357  
DB 612 PERLKEHD-----RIRLKAS 626

## RESULT 9

US-09-066-046-6  
Sequence 6, Application US/09066046A  
Patent No. 6204252  
GENERAL INFORMATION:  
APPLICANT: MURPHY, Cheryl  
STOREY, James  
BELTZ, Gerald A.  
COUGHLIN, Richard T.  
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTIC  
EHRILICHA AND METHODS OF USE  
NUMBER OF SEQUENCES: 65  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: HALE AND DORR LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: MA  
COUNTRY: United States  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/066,046A  
FILING DATE: 24-Apr-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Superko, Colleen  
REGISTRATION NUMBER: 39,850  
REFERENCE/DOCKET NUMBER: 106,941,155  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 526-6000  
TELEFAX: (617) 526-5000  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 578 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-066-046-6

Query Match 6.0%; Score 111.5; DB 3; Length 578;  
Best Local Similarity 18.6%; Pred. No. 0.0039;  
Matches 88; Conservative 77; Mismatches 172; Indels 137; Gaps 20;  
QY 7 IORRSTRIVGNVP-----IGDGA-----PIAVQSMNTRT-TDVEATVNO- 47

DB 112 VQEEBGTGYLLINAPKAVVRFKKIEKSAABEPQTVDPSVVSATSGSDVDTQEEGIDQE 171  
QY 48 IKALERVGADIVRSVPTMDAAEAPFLIQOVNVPVADIHFDYRLAKVAEYGVDCIAI 107  
DB 172 APAIEEVEETEEQEV---ILIEBGLTLDLEQPAVQPVVAEALP---GVAEAEIVPSL-- 223  
QY 108 NPGNIGNERIRMVV-----DCARDKNIPLR-----IGNVAGSLKDLQEKYGEPTP 154  
DB 224 -----EENKLOEVVAPPAQQLSAPESVAPAPQPESTVLGVAEGDLKSEVSVEANADVP 277  
QY 155 QALLSASMRHVDHLRLNDFQKVSV----- 180  
DB 278 QKEVISQOQEQEIAELBEGTEAPVEYKTEEVILKEDTLIDLEQPAVQPVVAEALPQV 337  
QY 181 -----KASDVLAIVESYRL-LAKQIDQPLH-----LGITTEAGARGSAVKA 221  
DB 338 EAAEAIVPSLEENKQLEVVVAPEAQQLSAPESVAPAPQPESTVLGVTE-----GDLKSE 391  
QY 222 IGL-----GLLSEGIQDTLRLVSLADPVEIKVGF-----ILKSIRSRGINFIA 269  
DB 392 VSEVADAGMQOEAQISD--QETQATEVEKVEVSVEKTEPEVILIEBGLTLDLEQPAVQ 449  
QY 270 CPTCSROEPDVIGTVNAL-----EQLLEDITTPMDVSIIGCV--VNGPEALVSTLGVGT 322  
DB 450 VPVVAEALPVEVAEALVPSLEENKQLEVVVAPEAQQLSAPESVAPAPQPESTVLGVTE 509  
QY 323 GNKKSGL-----YEDGVKRLDNNMDIDQLEARIKAKASQDLERRIDV 367  
DB 510 GDLKSEVSVEADAGMQOEAQISDQETQATEVEKVEVSVEADAGMQOEL--VDV 561

## RESULT 10

US-09-252-991A-30838  
Sequence 30838, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 30838  
LENGTH: 980  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30838

Query Match 5.7%; Score 106.5; DB 4; Length 980;  
Best Local Similarity 20.7%; Pred. No. 0.033;  
Matches 88; Conservative 65; Mismatches 155; Indels 113; Gaps 20;  
QY 16 YGVNVPIDGAPIAVQSMNTRTTDVEATVNOIKALERVGADIVRSVPTMDAAEAPFLI 75  
DB 434 FVDVYRLPQAPRSHRTLVADQVVDVGT-----EGIHFPVPARKTAQPDQAARSLGV- 487  
QY 76 KOQVNVPLVADIHF-----DYRI-----ALKVAEYGVDCIRPNGNIGNEERIRMVDC 124  
DB 488 --QVGIAVGVVEFEGRRLECRFVVDGEAQRVNO-----RVAPEGGAAGVHAELLVVV 539  
QY 125 ARDKNIPIRIGVNGSLEKDLQEKYGEPT---POLLES-----ANRHVDHLRLNFD 174  
DB 540 TAQAOPLP-----LATVLAELVGEPAFVQPVDAEGFAADVLVLPVAEQQYMAFG 591  
QY 175 QFKVSVKASDVFLAVESYRLAKQIDQPLHIGI-----TEAGARGSG 216  
DB 592 QGVVLPQGVAVATGIEA-----RVETALILIGIGLVLAAPGLBEGHPQIERSEAGTA-AD 644



QY 241 AADPVEIKVFDILKSLRISRGINFACPTCSROEFDVIGTVALLEQLEDDITPMV 300  
DB 419 AADPVEIKVFDILKSLRISRGINFACPTCSROEFDVIGTVALLEQLEDDITPMV 478  
QY 301 SIIGCVNNGEALVSTLGTGNGKSGLYEDGVRKDRLDNNMIDOLEARIPAKASOLD 360  
DB 479 AVIGCVNNGEALVSTLGTGNGKSGLYEDGVRKDRLDNNMIDOLEARIPAKASOLD 537  
QY 361 EA 362  
DB 538 EA 539

RESULT 6  
US-09-328-352-7906  
; Sequence 7906, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 7906  
; LENGTH: 378  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-7906

Query Match 64.0%; Score 1195; DB 4; Length 378;  
Best Local Similarity 63.8%; Pred. No. 6.7e-122; Indels 0; Gaps 0;  
Matches 231; Conservative 65; Mismatches 66; Indels 0; Gaps 0;

QY 6 PIORRSTRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPT 65  
DB 12 PIORRSTRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPT 71  
QY 66 MDAEAFKLIQOONVPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMVVDA 125  
DB 72 MEAAEAFKLIQOONVPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMVVDA 131  
QY 126 RDKNIPRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPT 185  
DB 132 RDKNIPRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPT 191  
QY 186 FLAVESYRLAKOIQDPLHIGTEAGARSGAVSAIGLILSEIGDITLRSIADPV 245  
DB 192 FLAVESYRLAKOIQDPLHIGTEAGARSGAVSAIGLILSEIGDITLRSIADPV 251  
QY 246 BEIKVGFILKSLRISRGINFACPTCSROEFDVIGTVALLEQLEDDITPMVDSIIGC 305  
DB 252 BEIKVGFILKSLRISRGINFACPTCSROEFDVIGTVALLEQLEDDITPMVDSIIGC 311  
QY 306 VVNGPFGALVSTLGTGNGKSGLYEDGVRKDRLDNNMIDOLEARIPAKASOLD 365  
DB 312 VVNGPFGALVSTLGTGNGKSGLYEDGVRKDRLDNNMIDOLEARIPAKASOLD 371  
QY 366 DV 367  
DB 372 EI 373

RESULT 7  
US-09-198-452A-389  
; Sequence 389, Application US/09198452A  
; Patent No. 6559294  
; GENERAL INFORMATION:  
; APPLICANT: Griffiths, R.  
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments  
; thereof and uses thereof, in particular for the diagnosis, prevention  
; and treatment of infection

FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-29  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 389  
LENGTH: 621  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-389

Query Match 25.2%; Score 470.5; DB 4; Length 621;  
Best Local Similarity 38.9%; Pred. No. 2.5e-42;  
Matches 111; Conservative 55; Mismatches 84; Indels 35; Gaps 5;

QY 9 RKXSTRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPTMA 68  
DB 20 RKXSTRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPTMA 79  
QY 69 AEAFLIKQOONVPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMVVDA 111  
DB 80 AEAFLIKQOONVPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMVVDA 138  
QY 112 -----IGNEERIRMVVDCARDKNIPIRIGVAGSLEKDOEKYGEPTPALLS 160  
DB 139 IYTESVAGSLRLLEKFAVLEKCRIGKAMRIGVHSGLSERIMQYGD-TIEGVAS 197  
QY 161 AMRHVDHLRINFDQKVSASDVFLAVESYRLAKOIQDPLHIGTEAGARSGAVSAIGL 215  
DB 198 AMRHVDHLRINFDQKVSASDVFLAVESYRLAKOIQDPLHIGTEAGARSGAVSAIGL 257  
QY 216 GAVKSAIGLILSEIGDITLRSIADPV 260  
DB 258 GAVKSAIGLILSEIGDITLRSIADPV 302

RESULT 8  
US-09-252-991A-29105  
; Sequence 29105, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196,136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29105  
; LENGTH: 1058  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29105

Query Match 6.0%; Score 112.5; DB 4; Length 1058;  
Best Local Similarity 21.5%; Pred. No. 0.0084;  
Matches 95; Conservative 50; Mismatches 140; Indels 157; Gaps 21;

QY 6 PIORRSTRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPT 65  
DB 252 PIORRSTRIYGVNPIGDPAPIAVOSMTNTRTDTVEATVNOIKALERVGADIVRSVPT 296  
QY 66 M-DAEAFKLIQOONVPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMVVDA 99  
DB 297 M-DAEAFKLIQOONVPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMVVDA 354  
QY 100 YGVDCIRINPGNIGNEERIRMVVDCARDKNIPIRIGVAGSLEKDOEKYGEPTPALLS 133  
DB 355 YGVDCIRINPGNIGNEERIRMVVDCARDKNIPIRIGVAGSLEKDOEKYGEPTPALLS 413

```
Db 1 MHESP1KRRKSTRIVGNVPIDGAP1AVQSMNTTRTTTVEATVROIQSLERVGVDIVR 60
Qy 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 120
Db 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRQ 120
Qy 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPQALLSARHVDLRLNFDQFVSV 180
Db 121 VVDSARHNIPIRIGVAGSLKEDLOEKYGEPTPEALVESAMRHVDLRLNFDQFVSV 180
Qy 181 KASDVFLAVESYRLAKQIDQPLHLGITTEAGARSQAVSAIGLGLLSGIGDTLRVSL 240
Db 181 KASDVFLAVESYRLAKQIDQPLHLGITTEAGARSQAVSAIGLGLLSGIGDTLRISL 240
Qy 241 AADPVEIKVGFDLKSLRIRSGINFIACPTCSROEFVIGTVNALQRLIEDIITPMDV 300
Db 241 AADPVEIKVGFDLKSLRIRSGINFIACPTCSROEFVIGTVNALQRLIEDIITPMDV 300
Qy 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVR-KDRLDNDMDIDLEARIKASQSL 359
Db 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVRKDERFDNDNIIDLEAKIRAKAAML 360
Qy 360 DE 361
Db 361 DE 362
```

RESULT 4  
US-09-170-187-5  
Sequence 5, Application US/09170187  
Patent No. 6383745

GENERAL INFORMATION:

APPLICANT: Rather, Philip N.

TITLE OF INVENTION: Methods For Screening For Anticicrobials

TITLE OF INVENTION: Utilizing aarc And Compositions Thereof

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/170,187

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/827,190

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Peter G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: CASE-02443

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 365 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-170-187-5

Query Match 87.0%; Score 1622.5; DB 4; Length 365;  
Best Local Similarity 88.4%; Pred. No. 1,2e-168;

```
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;
Qy 1 MHNPAP1ORRKSRIYVGNVPIDGAP1AVQSMNTTRTTTVEATVNOIKALERVGADIVR 60
Db 1 MHNPAP1ORRKSRIYVGNVPIDGAP1AVQSMNTTRTTTVEATVNOIKALERVGADIVR 60
Qy 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 120
Db 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRQ 120
Qy 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPQALLSARHVDLRLNFDQFVSV 180
Db 121 VVDSARHNIPIRIGVAGSLKEDLOEKYGEPTPEALVESAMRHVDLRLNFDQFVSV 180
Qy 181 KASDVFLAVESYRLAKQIDQPLHLGITTEAGARSQAVSAIGLGLLSGIGDTLRVSL 240
Db 181 KASDVFLAVESYRLAKQIDQPLHLGITTEAGARSQAVSAIGLGLLSGIGDTLRISL 240
Qy 241 AADPVEIKVGFDLKSLRIRSGINFIACPTCSROEFVIGTVNALQRLIEDIITPMDV 300
Db 241 AADPVEIKVGFDLKSLRIRSGINFIACPTCSROEFVIGTVNALQRLIEDIITPMDV 300
Qy 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVR-KDRLDNDMDIDLEARIKASQSL 359
Db 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVRKDERFDNDNIIDLEAKIRAKAAML 360
Qy 360 DE 361
Db 361 DE 362
```

RESULT 5  
US-09-252-991A-22991  
Sequence 22991, Application US/09252991A  
Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfeld et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196,136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 22991

LENGTH: 547

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22991

Query Match 73.6%; Score 1372.5; DB 4; Length 547;  
Best Local Similarity 73.8%; Pred. No. 4,9e-141;

Matches 267; Conservative 46; Mismatches 48; Indels 1; Gaps 1;

```
Qy 1 MHNPAP1ORRKSRIYVGNVPIDGAP1AVQSMNTTRTTTVEATVNOIKALERVGADIVR 60
Db 1 MHNPAP1ORRKSRIYVGNVPIDGAP1AVQSMNTTRTTTVEATVNOIKALERVGADIVR 60
Qy 179 IHSASPIIRKRSKRIWGNVPVGDAP1AVQSMNTTETCDVATVAQIRRLDAGADIVR 238
Db 179 IHSASPIIRKRSKRIWGNVPVGDAP1AVQSMNTTETCDVATVAQIRRLDAGADIVR 238
Qy 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 120
Db 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRV 120
Qy 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPQALLSARHVDLRLNFDQFVSV 180
Db 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPEALVESAMRHVDLRLNFDQFVSV 180
Qy 181 KASDVFLAVESYRLAKQIDQPLHLGITTEAGARSQAVSAIGLGLLSGIGDTLRVSL 240
Db 181 KASDVFLAVESYRLAKQIDQPLHLGITTEAGARSQAVSAIGLGLLSGIGDTLRISL 240
Qy 241 AADPVEIKVGFDLKSLRIRSGINFIACPTCSROEFVIGTVNALQRLIEDIITPMDV 300
Db 241 AADPVEIKVGFDLKSLRIRSGINFIACPTCSROEFVIGTVNALQRLIEDIITPMDV 300
Qy 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVR-KDRLDNDMDIDLEARIKASQSL 359
Db 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVRKDERFDNDNIIDLEAKIRAKAAML 360
Qy 360 DE 361
Db 361 DE 362
```

DB 61 VRVSVPTMDAAAEFKLIKQOVNPELVADHFDRIALKAAYGVDCFTLRINGNIGNEE 120  
QY 117 RIRMVDDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPOLLESAR--HVHDLRLNFD 174  
DB 121 RIRMVDDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPOLLESARFTHVHDLRLNFD 180  
QY 175 QFRVSVKASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVKSAT--GIGLILSEGI 232  
DB 181 QFRVSVKASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVKSATFTGLILSEGI 240  
QY 233 GDTLRVSLAADPVEEIKVGFIDILKSLIRSRGINFIACPTCSROEPDVI--GTVNALQOR 290  
DB 241 GDTLRVSLAADPVEEIKVGFIDILKSLIRSRGINFIACPTCSROEPDVI--GTVNALQOR 300  
QY 291 LEDIITPMDSIIGCVNPGGALVSTLGVGNGKSGLYEDGVKRDRLD--NNDMIDOL 348  
DB 301 LEDIITPMDSIIGCVNPGGALVSTLGVGNGKSGLYEDGVKRDRLDFTNNDMIDOL 360  
QY 349 EARIRAKASQLEARRIDVQOVER 372  
DB 361 EARIRAKASQLEARRIDVQOVER 384

## RESULT 2

US-09-170-187-6  
Sequence 6, Application US/09170187  
Patent No. 6383745  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/170,187  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/827,190  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 384 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-170-187-6

Query Match 96.5%; Score 1800; DB 4; Length 384;  
Best Local Similarity 96.9%; Pred. No. 4; 96-188;  
Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;  
QY 1 MHNQAPIQRKSTRIVGVNPIGIDGAPIAVQSWTNTTRTDDVEATV--NQIKALERVGADI 58

DB 1 MHNQAPIQRKSTRIVGVNPIGIDGAPIAVQSWTNTTRTDDVEATV--NQIKALERVGADI 60  
QY 59 VRVSVPTMDAAAEFKLIKQOVNPELVADHFDRIALKAAYGVDC--LRINGNIGNEE 116  
DB 61 VRVSVPTMDAAAEFKLIKQOVNPELVADHFDRIALKAAYGVDCFTLRINGNIGNEE 120  
QY 117 RIRMVDDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPOLLESAR--HVHDLRLNFD 174  
DB 121 RIRMVDDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPOLLESARFTHVHDLRLNFD 180  
QY 175 QFRVSVKASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVKSAT--GIGLILSEGI 232  
DB 181 QFRVSVKASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVKSATFTGLILSEGI 240  
QY 233 GDTLRVSLAADPVEEIKVGFIDILKSLIRSRGINFIACPTCSROEPDVI--GTVNALQOR 290  
DB 241 GDTLRVSLAADPVEEIKVGFIDILKSLIRSRGINFIACPTCSROEPDVI--GTVNALQOR 300  
QY 291 LEDIITPMDSIIGCVNPGGALVSTLGVGNGKSGLYEDGVKRDRLD--NNDMIDOL 348  
DB 301 LEDIITPMDSIIGCVNPGGALVSTLGVGNGKSGLYEDGVKRDRLDFTNNDMIDOL 360  
QY 349 EARIRAKASQLEARRIDVQOVER 372  
DB 361 EARIRAKASQLEARRIDVQOVER 384

## RESULT 3

US-08-827-190-5  
Sequence 5, Application US/08827190  
Patent No. 5858367  
GENERAL INFORMATION:  
APPLICANT: Rather, Philip N.  
TITLE OF INVENTION: Methods For Screening For Antimicrobials  
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Medlen & Carroll, LLP  
STREET: 220 Montgomery Street, Suite 2200  
CITY: San Francisco  
STATE: California  
COUNTRY: United States of America  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/827,190  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Carroll, Peter G.  
REGISTRATION NUMBER: 32,837  
REFERENCE/DOCKET NUMBER: CASE-02443  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 705-8410  
TELEFAX: (415) 397-8338  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-827-190-5

Query Match 87.0%; Score 1622.5; DB 2; Length 365;  
Best Local Similarity 88.4%; Pred. No. 1; 26-168;  
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;  
QY 1 MHNQAPIQRKSTRIVGVNPIGIDGAPIAVQSWTNTTRTDDVEATV--NQIKALERVGADI 60

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:14:33 ; Search time 21 Seconds  
(without alignments)  
749.507 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAPIORRSTRIYGVN.....RAKASQDEARRIDVQVEK 372

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Database: Issued Patents\_Aa:\*

1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:.\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:.\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:.\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:.\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS\_COMB.pep:.\*  
6: /cgn2\_6/prodata/1/1aa/backfile1.pep:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1800	96.5	384	2	US-08-827-190-6
2	1800	96.5	384	4	US-09-170-187-6
3	1622.5	87.0	365	2	US-08-827-190-5
4	1622.5	87.0	365	4	US-09-170-187-5
5	1372.5	73.6	547	4	US-09-328-352-7906
6	1195	64.0	378	4	US-09-328-352-7906
7	470.5	25.2	621	4	US-09-198-452A-389
8	112.5	6.0	1058	4	US-09-252-991A-29105
9	111.5	6.0	578	3	US-09-066-046-6
10	106.5	5.7	980	4	US-09-252-991A-30838
11	105.5	5.7	578	3	US-08-975-762-50
12	105.5	5.7	578	3	US-09-295-028-50
13	105.5	5.7	578	3	US-09-295-028-50
14	101	5.4	465	4	US-09-106-582-50
15	101	5.4	465	4	US-09-252-991A-29387
16	100.5	5.4	620	4	US-09-328-352-7730
17	100	5.4	343	3	US-09-328-352-7730
18	99.5	5.3	343	3	US-09-328-352-7730
19	97	5.2	1037	4	US-09-252-991A-24636
20	96.5	5.2	851	4	US-09-134-001C-4794
21	96.5	5.2	851	4	US-09-134-001C-4794
22	95.5	5.1	303	4	US-09-252-991A-24773
23	95.5	5.1	740	1	US-08-309-512-10
24	95.5	5.1	740	5	PCT-US92-08756A-10
25	95	5.1	796	4	US-09-252-991A-17763
26	94.5	5.1	483	4	US-08-887-534A-51
27	94.5	5.1	483	4	US-09-527-431-51

28	94.5	5.1	659	4	US-09-252-991A-17904	Sequence 17904, A
29	94.5	5.1	1503	3	US-08-976-255-14	Sequence 14, Appl
30	93.5	5.0	619	4	US-09-252-991A-17411	Sequence 17411, A
31	93.5	5.0	659	4	US-09-252-991A-17731	Sequence 17731, A
32	93	5.0	459	4	US-09-491-785-2	Sequence 2, Appl1
33	93	5.0	483	4	US-09-252-991A-27988	Sequence 27988, A
34	93	5.0	607	4	US-09-252-991A-20596	Sequence 20596, A
35	92.5	5.0	489	4	US-09-252-991A-27651	Sequence 27651, A
36	92.5	5.0	940	4	US-09-512-250C-2	Sequence 2, Appl1
37	91	4.9	470	4	US-09-252-991A-26312	Sequence 26312, A
38	91	4.9	817	4	US-09-252-991A-31147	Sequence 31147, A
39	91	4.9	1181	4	US-09-252-991A-18480	Sequence 18480, A
40	90.5	4.8	484	4	US-09-252-991A-20787	Sequence 20787, A
41	90.5	4.8	489	4	US-09-252-991A-19810	Sequence 19810, A
42	90.5	4.8	573	4	US-09-252-991A-26428	Sequence 26428, A
43	90.5	4.8	1612	1	US-08-169-927-2	Sequence 2, Appl1
44	90	4.8	358	4	US-09-784-508-4	Sequence 4, Appl1
45	90	4.8	629	4	US-09-252-991A-31575-	Sequence 31575, A

## ALIGNMENTS

RESULT 1  
US-08-827-190-6  
; Sequence 6, Application US/08827190  
; Patent No. 5858367  
; GENERAL INFORMATION:  
; APPLICANT: Rather, Philip N.  
; TITLE OF INVENTION: Methods For Screening For Antimicrobials  
; TITLE OF INVENTION: Utilizing aarc And Compositions Thereof  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Medien & Carroll, LLP  
; STREET: 220 Montgomery Street, Suite 2200  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: United States of America  
; ZIP: 94104  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/827,190  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Carroll, Peter G.  
; REGISTRATION NUMBER: 32,837  
; REFERENCE/DOCKET NUMBER: CASE-02443  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 705-8410  
; TELEFAX: (415) 397-8338  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 384 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-827-190-6  
Query Match 96.5%; Score 1800; DB 2; Length 384;  
Best Local Similarity 96.9%; Pred. No. 4.9e-188;  
Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;  
QY 1 MHNQAPIORRSTRIYGVNPIGDCAPIAVOSMTNRTTVEATV--NQIKALEVGVADI 58  
DB 1 MHNQAPIORRSTRIYGVNPIGDCAPIAVOSMTNRTTVEATVFTNQIKALEVGVADI 60  
QY 59 VRVSPFTMDADEFILINQGVNVPVADIHFDYRIALAKVAEYGVDC--LRINPGINIGE 116

**THIS PAGE BLANK (uspto)**

```

Db      262 FEFAR1CRKLDYHNFVFSMKA5NPVIMVQAYRLVAVEMVHGWDYPLHLGVTGAGEGEGD 321
Qy      217 AVKSAIGLLLSGIGDTRLVSLADPVEI 248
Db      322 RMKSAIGITLQDGLDITIRVSLTEPPEEI 353

```

## RESULT 15

```

08KG23 PRELIMINARY; PRT; 746 AA.
ID 08KG23
AC 08KG23;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE GCPE proteain.
GN GCPE OR CTO147.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxId=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Patksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
RA Vamathavan J., Khouli H., White O., Gruber T.M., Ketchum K.A.,
RT "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.",
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL; AE012794; AAM71395.1; -.
DR TIGR; CTO147; -.
DR InterPro; IPR006705; GCPE.
DR InterPro; IPR004588; IspG.
DR Pfam; PF04551; GCPE; 2.
DR TIGRFAMs; TIGR00612; IspG_gcPE; 1.
KW Complete proteome.
SQ SEQUENCE 746 AA; 82087 MW; B40BD95B6010C5F9 CRC64;

```

```

Query Match      24.4%; Score 455.5; DB 16; Length 746;
Best Local Similarity 37.5%; Pred. No. 3.8e-25;
Matches 108; Conservative 51; Mismatches 90; Indels 39; Gaps 8;

```

```

Qy      5 API---QRRKSTRIVYGVNPIGDGAPIAVQSMTRTRTTVEATVNOIKALERYGADIVRV 61
Db      68 APVYSYRRRVTRVPPGTIFLGGYLPFRVSMITAHMTDPAASVEQCRRLYEAGCEIIRL 127
Qy      62 SVPTMDAAEAPFKLIKQO-----VNPPLVADIHFDYRIALKVAEYGVDCRLINPNTGN-- 114
Db      128 TVPTKDAENLKNIRQLRRDGDIDTLPVADIHFSAKAMKAVEF--VENIRINPGNATGA 186
Qy      115 -----EERIRMVDCARDKNIPIRIGVNVAGSLKXQLEKYEGETPQ 155
Db      187 KSSSKDYTDDEYRAELDKVREBFTPLVKRKSIGVSMRIGTNHGSLSDRIVSRYN-SPB 245
Qy      156 ALLESAMRHVDHL-DRLNFDQFKSVKASDVFLAVESYRLAKOID-----QPLHLGIT 208
Db      246 GWVEALLESFRICEDEGYDQL-FSMKSSNVVRMIGAYRLVLVARADAEIRYAFPLHLGVT 304
Qy      209 EAGGARSGAVKSAIGLLLSGIGDTRLVSLADPVEEIKVGFILK 256
Db      305 EAGDGDGEGRIKSAIGLLLSGIGDITIRVSLTEPVEVPVGFALVK 352

```

Search completed: November 23, 2003, 17:16:38  
 Job time : 63 secs

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Gcpe protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Columbia;  
 RX MEDLINE=97471969; PubMed=9330910;  
 RA Sato S., Kocani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,  
 RA Miyajima N., Tabata S.;  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence  
 RT features of the 1.6 Mb regions covered by twenty physically assigned  
 RT clones."  
 RL DNA Res. 4:215-230(1997).  
 DR EMBL; AB005246; BAB09833.1; -  
 DR InterPro; IPR004588; IspG.  
 DR TIGRFBMS; TIGR00612; IspG gcpe; 1.  
 SQ SEQUENCE 716 AA; 79817 MW; 28D36FC64EB7CA CRC64;

Query Match 24.8%; Score 462; DB 10; Length 716;  
 Best Local Similarity 38.6%; Pred. No. 1.2e-25;  
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RKRSTRIYVGNVPIGAGPIAVQSMNTRTTVEATVNOIKALERVADIVRSVPTMDA 68  
 DB 59 RKRTRTVMGVNLGSEHPIRIGTMTSDTKDITGVDEVMRIADKADIVRTVOGKKE 118  
 QY AEAFA-----KLTKQVNVPLVADIHFDYRIALKVAEYVDCRLINPNGINNEERIRMVVD 123  
 DB 119 ADACFEIKKLVQNLNINIPVADIHFAPTVALRVAC-FDKIRVNGNFPADRAQFETID 177  
 QY 124 CARDK-----NIPRIGVNASGLEKDLQEKYGEPTPQALLESA 161  
 DB 178 YTEDEVQKELQHIIEQVFTPLVEKCKKYGAMRIGTNHGSLSDRIMSYGD-SPRGWESA 236  
 QY 162 MRHVHDLRLNFDQFVSVKASDVFLAVESYRLAKOI-----DQPLHIGITEAGARG 216  
 DB 237 FEFARICRKLVDYHNFEVSMKASNPVIMVQAYRLVAVMGWDYPLHIGITEAGGEDG 296  
 QY 217 AVKSAIGLILLSEGIQDITRVSIAADPVEEI 248  
 DB 297 RKMSAIGITLQDGLDGLTIRVSLTEPPEEII 328

## RESULT 13

ID 08GZ87 PRELIMINARY; PRT; 740 AA.  
 AC 08GZ87;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE Gcpe.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Querol J., Campos N., Imperial S., Boronat A.,  
 RA Rodriguez-Concepcion M.;  
 RT "Identification and functional analysis of plant orthologs of the  
 RT Escherichia coli gcpE gene product with a role in plastid isoprenoid  
 RT biosynthesis."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF344673; AA015446.1; -  
 SQ SEQUENCE 740 AA; 82157 MW; BEF625E9A9C88074 CRC64;

Query Match 24.8%; Score 462; DB 10; Length 740;

Best Local Similarity 38.6%; Pred. No. 1.2e-25;  
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RKRSTRIYVGNVPIGAGPIAVQSMNTRTTVEATVNOIKALERVADIVRSVPTMDA 68  
 DB 84 RKRTRTVMGVNLGSEHPIRIGTMTSDTKDITGVDEVMRIADKADIVRTVOGKKE 143  
 QY AEAFA-----KLTKQVNVPLVADIHFDYRIALKVAEYVDCRLINPNGINNEERIRMVVD 123  
 DB 144 ADACFEIKKLVQNLNINIPVADIHFAPTVALRVAC-FDKIRVNGNFPADRAQFETID 202  
 QY 124 CARDK-----NIPRIGVNASGLEKDLQEKYGEPTPQALLESA 161  
 DB 203 YTEDEVQKELQHIIEQVFTPLVEKCKKYGAMRIGTNHGSLSDRIMSYGD-SPRGWESA 261  
 QY 162 MRHVHDLRLNFDQFVSVKASDVFLAVESYRLAKOI-----DQPLHIGITEAGARG 216  
 DB 262 FEFARICRKLVDYHNFEVSMKASNPVIMVQAYRLVAVMGWDYPLHIGITEAGGEDG 321  
 QY 217 AVKSAIGLILLSEGIQDITRVSIAADPVEEI 248  
 DB 322 RKMSAIGITLQDGLDGLTIRVSLTEPPEEII 353

## RESULT 14

ID 08LPO4 PRELIMINARY; PRT; 741 AA.

AC 08LPO4;  
 DT 01-OCT-2002 (Tremblrel. 22, Created)  
 DT 01-OCT-2002 (Tremblrel. 22, last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)  
 DE AT5G60600/mup24.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 NC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,  
 RA Bowser L., Carroll P., Chang E., Dale J.M., Goldsmith A.D.,  
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,  
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,  
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,  
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,  
 RA Becker J.R.;  
 RT "Arabidopsis cDNA clones."  
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY094472; AAM19840.1; -  
 DR InterPro; IPR006705; Gcpe.  
 DR InterPro; IPR004588; IspG.  
 DR Pfam; PF04551; Gcpe; 2.  
 DR TIGRFBMS; TIGR00612; IspG gcpe; 1.  
 SQ SEQUENCE 741 AA; 82257 MW; BCALD3147BD63ACB CRC64;

Query Match 24.5%; Score 457; DB 10; Length 741;

Best Local Similarity 38.2%; Pred. No. 2.9e-25;

Matches 104; Conservative 48; Mismatches 86; Indels 34; Gaps 5;

QY 9 RKRSTRIYVGNVPIGAGPIAVQSMNTRTTVEATVNOIKALERVADIVRSVPTMDA 68  
 DB 84 RKRTRTVMGVNLGSEHPIRIGTMTSDTKDITGVDEVMRIADKADIVRTVOGKKE 143  
 QY AEAFA-----KLTKQVNVPLVADIHFDYRIALKVAEYVDCRLINPNGINNEERIRMVVD 123  
 DB 144 ADACFEIKKLVQNLNINIPVADIHFAPTVALRVAC-FDKIRVNGNFPADRAQFETID 202  
 QY 124 CARDK-----NIPRIGVNASGLEKDLQEKYGEPTPQALLESA 161  
 DB 203 YTEDEVQKELQHIIEQVFTPLVEKCKKYGAMRIGTNHGSLSDRIMSYGD-SPRGWESA 261  
 QY 162 MRHVHDLRLNFDQFVSVKASDVFLAVESYRLAKOI-----DQPLHIGITEAGARG 216

DR EMBL; A014468; AAN30676.1; --  
 KW TIGR; BR178; --  
 KM Complete proteome.  
 SQ SEQUENCE 420 AA; 45029 MW; F1495DB2D73E164C CRC64;

Query Match 27.2%; Score 508; DB 16; Length 420;  
 Best Local Similarity 33.2%; Pred. No. 2.2e-29;  
 Matches 128; Conservative 79; Mismatches 109; Indels 70; Gaps 13;

QY 6 PIGRRSTRIVGNVPIDGAPVAVQSMNTRTTVDVATVNOIKALERVGADIVRSVPT 65  
 DB 12 PFRPRGSGVSGVGVVGSAPVAVQSMNTTADVSTVAGVAAHRAAGSEIVRTVDR 71  
 QY 66 MDAEAFKLIQOV-----NVPLVADIF-----DY-RIALKVAEYGVDCRLINPNI 112  
 DB 72 DESAAAVPKIRERLERGHVPLVGPDPHYIGHKLADHPACAEALAKY-----RINPNI 126  
 QY 113 G----NEERIRMVVDCARDKNIPIRIGVNAAGLEKDI-----QEKYGEPT-----Q 155  
 DB 127 GFKDKDKQPADIVEMAIRYDKKVRIGVNMGSIDQELTLTMDRMQAEAPLSADVMRE 186  
 QY 156 ALLESAMRHVDHLDRLNF--DQKVSVKASDVFLAVESYLLAKQIDQPLHIGITAGGA 213  
 DB 187 AIVQSAALISANLAEIIGLGRDKITLSAKVQVODLIAVYMLQKRSNHLHLGLTENGWG 246  
 QY 214 RSGAVKSAIGLLSEGIGDTLRVSLAADP-----VEEIKVGFPIKSLIRSRGINFIA 269  
 DB 247 TKGIVASSAMGILLQGGIGDTIRISLTPEPGDRIREVQVAGELQTMFRQFPIVAA 306  
 QY 270 CPTCSQOEPFVIGTVALAEQRL-----EDI-----ITPVDISITCCVNGPG 311  
 DB 307 CFCGCR-----TTSVFQELAQTIQEDIRNMPLREKYPGEVALSVVMGCIYNGPG 359  
 QY 312 EALVSTLGV-----TGKMKSGLYEDG 333  
 DB 360 ESKHADIGISLPGETGPPSAPVVDG 385

## RESULT 10

Q8GZR6 PRELIMINARY; PRT; 740 AA.  
 AC Q8GZR6;  
 DT 01-MAR-2003 (Tremblrel. 23, Created)  
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE GCPE.  
 OS Lycopersicon esculentum (Tomato).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 OC Asteridae; lamids; Solanales; Solanaceae; Solanum.  
 OC NCBI\_Taxid=4081;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Querol J., Imperial S., Boronat A., Rodriguez-Concepcion M.;  
 RT "Identification and functional analysis of plant orthologs of the  
 RT Escherichia coli gcpe gene product with a role in plastid isoprenoid  
 RT biosynthesis";  
 RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF435086; AAO15447.1; --  
 SQ SEQUENCE 740 AA; 82230 MW; 8668C9F08078FA05 CRC64;

Query Match 25.2%; Score 471; DB 10; Length 740;  
 Best Local Similarity 40.1%; Pred. No. 2.7e-26;  
 Matches 109; Conservative 45; Mismatches 84; Indels 34; Gaps 5;

QY 9 RKRSTRIVGNVPIDGAPVAVQSMNTRTTVDVATVNOIKALERVGADIVRSVPTMDA 68  
 DB 84 RRGTCVTVMGNVAGSEHPIRIGTMTTDDTKDVAALVEGQMKADAGADIVRTVQGRKE 143  
 QY 69 AEAFF-----KLKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPNI----- 114  
 DB 144 ADACFEIKQTLVQKNYINPLVADIHFAFPAVALRAEC-FDKIRVNGNPNADRRAPPEOLE 202

QY 115 -----EERIRMVVDCARDKNIPIRIGVNAAGLEKDIQEKYGEPTPOALLESA 161  
 DB 203 YTEDDYQKELEHIEVEFTPLVEKCKYGRAMRIGTNHGLSIRIMSYGD-SPRGWESA 261  
 QY 162 MRHVHDLRLNPDQFVSVKASDVFLAVESYRLAKQI-----DQPLHIGITAGGARSG 216  
 DB 262 FEFARICRLDHFNFVFSMKASNPVVMQAYRLVAEMVYGVGDYPLHIGITEAGGEDG 321  
 QY 217 AVKSAIGLLSEGIGDTLRVSLAADPVEEI 248  
 DB 322 RMKSAIGITGLQDGLGDTIRVSLTEAPEEI 353

## RESULT 11

Q8RXG8 PRELIMINARY; PRT; 741 AA.  
 AC Q8RXG8;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)  
 DE GCPE protein.  
 GN AF560600.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_Taxid=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carinci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY081261; AAL91150.1; --  
 DR InterPro: IPR006705; GCPE.  
 DR InterPro: IPR004588; IspG.  
 DR Pfam: PF04551; GCPE; 1.  
 DR TIGR: TIGR00612; IspG\_gcpe; 1.  
 SQ SEQUENCE 741 AA; 82157 MW; 957B11538F481585 CRC64;

Query Match 24.8%; Score 463; DB 10; Length 741;  
 Best Local Similarity 38.6%; Pred. No. 1e-25;  
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RKRSTRIVGNVPIDGAPVAVQSMNTRTTVDVATVNOIKALERVGADIVRSVPTMDA 68  
 DB 84 RKRTRVVMGNVAGSEHPIRIGTMTTSDTKDITGTVDEVMRIADKAGADIVRTVQGRKE 143  
 QY 69 AEAFF-----KLKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPNI----- 123  
 DB 144 ADACFEIKQTLVQKNYINPLVADIHFAFPAVALRAEC-FDKIRVNGNPNADRRAPPEITD 202  
 QY 124 CARDK-----NIPRIGVNAAGLEKDIQEKYGEPTPOALLESA 161  
 DB 203 YTEDEYQKELEHIEVEFTPLVEKCKYGRAMRIGTNHGLSIRIMSYGD-SPRGWESA 261  
 QY 162 MRHVHDLRLNPDQFVSVKASDVFLAVESYRLAKQI-----DQPLHIGITAGGARSG 216  
 DB 262 FEFARICRLDHFNFVFSMKASNPVVMQAYRLVAEMVYGVGDYPLHIGITEAGGEDG 321  
 QY 217 AVKSAIGLLSEGIGDTLRVSLAADPVEEI 248  
 DB 322 RMKSAIGITGLQDGLGDTIRVSLTEAPEEI 353

## RESULT 12

Q9FF59 PRELIMINARY; PRT; 716 AA.  
 AC Q9FF59;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)



```

QY 189 VESYRLANKIDPRLHIGITEAGASGVKSAIGLGLLSEIGDITLRSVLADPVEE 248
DB 185 IEAVYKAAAEFNYPLHIGITEAGSLFENGITKSAAGVLLHIEGINTIRISLGGDPLSEV 244
QY 249 KVFQDLKSLRISRGINFACPTCSROEPDVIGTVALLBQRLIEDITTPMDVSIICVYN 308
DB 245 KVGCKLLNSIGLVLDNLVDSICTPGLNDLNPVYKEIFKTRKNFPLKVAIILGCAYN 304
QY 309 GPGFALVSTLGVTVGNKSGLYEDGVKRDLDNNMDIDQLEARIKAKASQLEARR 364
DB 305 GPGFAKEADIGIAGNGTGIIFANNGKAIKSVPEQDVLDELKILISIKYKEYLDSKK 360

RESULT 7
Q8DK70 PRELIMINARY; PRT; 402 AA.
Q8DK70;
AC 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE TLR0996.
GN TLR0996.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_Taxid=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=2225144; PubMed=12240834;
RA Nakamura Y., Kaneo T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Igaruchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002)
DR EMBL; AP005372; BAC08548.1; -
KW Complete proteome.
SQ SEQUENCE 402 AA; 44192 MW; 7595FFC940AC60D7 CRC64;

Query Match 32.2%; Score 601.5; DB 16; Length 402;
Best Local Similarity 36.8%; Pred. No. 2,4e-36;
Matches 147; Conservative 79; Mismatches 120; Indels 53; Gaps 11;

QY 4 QADIQRRKSTRIVYGNVPIGDGAPIAVQSMTRTTDVEATVNOIKALEVAGADIVRSV 63
DB 14 ETAVIRKTRPPIGVSIVIGGHPVAVQSMINEDTLDIEGVAIRRLHEIGCEIVATV 73
QY 64 PTMDAAEAF-----KLIKQVNPVLVDIHPD-YRIKLKAEYGVDCRLINPG----- 110
DB 74 PSLAAKAMEIRDLRYKYPVPLVDVHNGKIKALEVAKY-VDVVRINPGLYVEKP 132
QY 111 -----NIGNE--ERIRVVDCARDKNIPIRIGVNGSLSEKDOEKYGEPTPQ 155
DB 133 KPNRTETQAFDEIGAKIKETLEPLVISLRDQSKSRIGVNGSLAERMLFTYGD-TPE 191
QY 156 ALLESAMRHVDHLDRLNFPDQFKVSVKASDVFLAVESYRLAKQIDQ-----PLHLGITEA 210
DB 192 GWVESALEFIRICSLNFYMLISLTKASRYVPMIAANRLMKRMDELGMYPPLHLGITEA 251
QY 211 GGARSGAVKSAIGLGLLSEIGDITLRSVLADPVEEIKYGFDLKSLRISGINFIAC 270
DB 252 GDGEYGIKSTAGIATLALRIGITIVSLTEAPEKEIPCYGLILGLRRTWVEVAC 311
QY 271 PTCRSQFVDVIGTVNALEQRLEDI-----ITPMDSIICVNGPPEALVSTLGVTVGN 324
DB 312 PSCGRTIFN-----LEEVIAHKREARKILTGINIAVMGIVNGPEMADADGVYGV-- 362
QY 325 KKSQ--LYEDGVKRDLDNNMDIDQLEARIKAKASQLE 360
DB 363 KQPGYISLYRGREVRKVPPEAEGVAALVELIKADGRWVD 401

```

```

RESULT 8
Q8FLH5 PRELIMINARY; PRT; 663 AA.
Q8FLH5;
AC 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE GCPE protein homolog.
GN GCPE OR LA3160.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_Taxid=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011478; AA050358.1; -
KW Complete proteome.
SQ SEQUENCE 663 AA; 73779 MW; 79B60DF805A6365 CRC64;

Query Match 28.2%; Score 525.5; DB 16; Length 663;
Best Local Similarity 41.7%; Pred. No. 2.1e-30;
Matches 118; Conservative 53; Mismatches 75; Indels 37; Gaps 8;

QY 2 HNOAPI--QRKSTRIVYGNVPIGDGAPIAVQSMTRTTDVEATVNOIKALEVAGADIV 59
DB 5 YNOTPEGYORRRRTREYKVDGVKGNPNPIVQSMINSDTDTQGSVKQILEERAGCEIV 64
QY 60 RVSVPMDAAEAFKLIKQOV-----NVPLVADIHFPYRIALKAEXGVDCRLINPGIN 114
DB 65 RLIVPQADADNLPSIRQELKAKGSRVPLVDIHFPYSAMKAVEI-VEKVRINPGENFD 123
QY 115 E-----ERIR-----MVVDCARDKNIPIRIGVNGSLSEKDOEKYGEPTPQ 152
DB 124 KKKFAVRDYTDLEYNQELERISEVFPVLVRC-KELGVSMRIGTNGHSLDRIMRYGD- 181
QY 153 TPQALLSARHVDHLDRLNFPDQFKVSVKASDVFLAVESYRLAK-----QIDQPLHLGI 207
DB 182 TPGWVESALEFIRIAESLGYDYDIIVSMKASNPQVAVQYRLMASFNEELKMDYPLHLGV 241
QY 208 TEAGARSGAVKSAIGLGLLSEIGDITLRSVLADPVEEIKY 250
DB 242 TEAGDNGDRIRKSAIGSLDEGLDITRVSLETPVLEVEY 264

RESULT 9
Q8FYT2 PRELIMINARY; PRT; 420 AA.
Q8FYT2;
AC 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE GCPE protein.
GN GCPE OR BR1778.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_Taxid=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Debroy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmiller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Frazer C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).

```

**ISFG OR BL0098.**

OS Bifidobacterium longum.  
OC Bacteria; Actinobacteria; Actinobacteridae; Bifidobacterales;  
NC Bifidobacteriaceae; Bifidobacterium.  
OX NCBI\_TaxID=216816;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NCC 2705;  
RX MEDLINE=22294977; PubMed=12381787;  
RA Schell J.M., Karmaliantzou M., Snel B., Vilanova D., Berger B.,  
R Pesel G., Zwielen M.-C., Desiere F., Bork P., Delley M.,  
RA Pidmore R.D., Arigoni F.;  
RT "The genome sequence of Bifidobacterium longum reflects its adaptation  
to the human gastrointestinal tract." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).  
DR EMBL; AE014624; AAN23963.1; - .

KM Complete proteome.  
SQ SEQUENCE 403 AA; 42564 MW; FF6F8CAJ615A6O5E CRC64;

Query Match 43.6%; Score 814.5; DB 16; Length 403;  
Best Local Similarity 46.3%; Pred. No. 4.2e-52;  
Matches 167; Conservative 70; Mismatches 111; Indels 13; Gaps 4,

DQ 3 NOAPIQ-RRKSTRIYVGNVPILGDPAPIAVSMWNTNRTPDVEATVNQIKALERVGVDIRV 61  
::.:|:||||| ||||| |.|.|||.||||| :|:|:|:|:  
DB 22 SESPLHPRKRKRRIRIIVGPVGGGAPISVSQTNLTLNVPTLOQLIELTAAAGCDIVRV 81  
  
DQ 62 SVPTMDAAEAEEKLIKQVNVLPVLVDIHFDYRIALKVAEGVDCLRINPGINIGERIRM- 120  
::.:|:||||| ||||| |.|.|||.||||| :|:~::~|:  
DB 82 AVPEQQDDADALPEICRSPIVIADIHQSKKYFPQADAGCAAAYVNGNIKKPVEVPD 141  
::.:|:||||| ||||| |.|.|||.||||| :|:~::~|:  
  
DQ 121 VVDCARDKNIPDIRIGNAGSLSEKDLOEKYEETPOLLESAMRHVDHLDRINFDFKVSV 180  
142 ICKATPDAGILRGVNNAGSLDKELPYAKYGPTPALVASLKAEAMFEDEVGFHFXTSV 201  
  
DQ 181 KASVFLLAVESYRLIAKOIDDPHGHTEAGAARGSAVKSAIGCLLSSEGDTLRSL 240  
::.:|:||||| ||||| |.|.|||.||||| :|:~::~|:  
DB 202 KHHDVTIMWEYTRILLASKDWPLHGLGVEADPAWGTTKSCLAIFALLAEEGTITRVSL 261  
  
DQ 241 AADVEERIIEFIDLKSRISRGINFACTPSROEPDIVGITVALLEQRLEDITTPMDV 300  
262 SAPPAEEVRKVCCKLEYMGLPKRKDIITSCSGCGAQADVQLASAVEGLEKDYAFARV 321  
  
DQ 301 SIICVVNPGEALVSTLGTVTGNNKGSLVEDG-----VRDRLDN-----NDMIQLE 349  
322 AMGCIIVNGPEARBADLVGASGNGKGIPTFKGVKITVPEDQIVDTLLTIANDIAAME 381  
350 A 350  
382 A 382

ID Q8FP82 PRELIMINARY; PRT: 393 AA.  
AC Q8FP82;  
DT 01-MAR-2003 (TREMBLrel\_23, Created)  
DT 01-MAR-2003 (TREMBLrel\_23, Last sequence update)  
DT 01-MAR-2003 (TREMBLrel\_23, Last annotation update)  
DE Putative aminoglycoside acetyltransferase negative regulator.  
CN CE1903,  
OS Corynebacterium efficiens.  
OC Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
CX NCBI\_TaxID=152794;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Y5-314 / AJ 12310 / DSM 44549 / JCM 11189;  
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,  
RA Ikeo K., Suzuki M., Maehima J., Itoh T., Yamagishi A., Nishio Y.,  
RA Usuda Y., Sugimoto S.;  
RT "The entire genomic sequence of Corynebacterium efficiens Y5-314.",  
BL Submitted (May-2002) to the EMBL/GeneBank/DBJ databases.

DR	EMBL; AP005220; BAC18713.1; --					
KM	Transferase; Complete proteome.					
SQ	SEQUENCE 393 AA; 41533 MW; A8CC2A9BF150AD66 CRC64;					
Query Match	40.4%; Score 753; DB 16; Length 393;					
Best Local Similarity	43.7%; Pred. No. 1.4e-47;					
Matches	157; Conservative 67; Mismatches 123; Indels 12; Gaps 3					
OY	9	RRKSTRIVGVNPIGDGAPLAVQSMTWTRTDDVEATVNOIKALERVGADIVRSVPTMDA	68			
DB	20	RRKTRQLMWGSVGVSGDHPISVSQSTTKTHDHINATLIQQIAGLTASGCDIVRACPKEVD	79			
OY	69	AEAFLKIQQOVNVPLVADIHFDYRIALKAVEYGVDCILRNFGNIGNEERIRMVVDCARD	127			
DB	80	AEALPFIAKKSPDIVADIHFPOPKFTSADIDAGCAAAVRVNGNIKEFGRVLEVAQAAGD	139			
OY	128	KNIPFIRGVANGLSEKLQGEKY-GEPFPQALLSEAMRHVDHLDRINFQPFVSVASDVDF	186			
DB	140	AGIFPRIGVNGSLDKRLIDRKHGAIPEALVESLWEASLPFEHGYGDIAISVGHSDPV	199			
OY	187	LAVESYRLLAQIQOPLHTGITTEAGARGASGAVKSAILGLLSLEGIDTLRVSLADAVE	246			
DB	200	LMVEAYRQLABKCQYPHLGTVEAGRPKMGTIKSVAFGALLSGIGITIVSLADAVE	259			
OY	247	EIKYGFDIKSLRIRSRCINFIACTPSGRDPYIGYNALQRLEDITTPMDYSIIICV	306			
DB	260	EIKYGDOITLOSLNRPRKLETIVSCPCGRAQVDYKLAEEVTGEDGLEVLBRVAVMGCV	319			
OY	307	VNGGEMALVSTLGVTGNKKSGLEVDPGRKORLDNNIMIDLPEARIRAKASQDLBARRI	365			
DB	320	VNGGEGARADADLVASGNGKQITVKGEIITKTPESPQIVQTL-----IEEAMRI	368			
RESULT 6						
O8EU16						
ID	O8EU16	PRT:	362 AA.			
AC	O8EU16:	PRELIMINARY;				
DT	01-MAR-2003 (TREMBLrel. 23, Created)					
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)					
DE	01-MAR-2003 (TREMBLrel. 23, Last annotation update)					
DE	Peptidoglycan acetylation.					
GN	MYPE9400.					
OS	Mycoplasma penetrans.					
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.					
OX	NCHI_TaxID=28227;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=HF-2;					
FX	MEDLINE=22354719; PubMed=12466555;					
RA	Sasaki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K., Yoshino C., Horino A., Shibata T., Sasaki T., Hattori M.;					
RT	"The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans.";					
RL	Nucleic Acids Res. 30:5293-5300(2002).					
DR	EMBL; AP004174; BAC44727.1; -.					
KM	Complete proteome.					
SQ	SEQUENCE 362 AA; 39237 MW; 3FEB12EF331AE73 CRC64;					
Query Match	39.6%; Score 739.5; DB 16; Length 362;					
Best Local Similarity	41.0%; Pred. No. 1.3e-46;					
Matches	146; Conservative 82; Mismatches 127; Indels 1; Gaps 1;					
OY	9	RRKSTRIVGVNPIGDGAPLAVQSMTWTRTDDVEATVNOIKALERVGADIVRSVPTMDA	68			
DB	6	REKTKAVYGVNQIGGNKKVVIQSMITTKTHDYEKVTAQVKEYLREGCELYRIAVLDDDED	65			
OY	69	AEAFLKIQQOVNVPLVADIHFDYRIALKAVEYGVDCILRNFGNIGNEERIRMVVDCARD	128			
DB	66	AAARGEVVVNSPCPIIADIHFNPLVYALKAESGAAYKRLNPGNIKDEQJRKIIDLANKK	125			
OY	129	NIPRIGVANGLSEKLDQEKYGEPTPOLLESAMRHVDHLDRINFDPFKYSVKASDVFLA	188			
DB	126	NIPRIVGVNGSLFPMDLMKSHG-VTADAMMIAVRGYINLEESNGFNFIIVSIKATNVLLA	184			

QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVKSATGILLSGIGDTLRVSL 240  
 DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVKSATGILLSGIGDTLRVSL 240  
 QY 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMV 300  
 DB 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMV 300  
 QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKAKASQL 359  
 DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKAKASQL 359  
 QY 360 DEARIDVQ 368  
 DB 361 DESNRIDIK 369

## RESULT 2

Q8EC32 PRELIMINARY; PRT; 371 AA.  
 AC Q8EC32;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISP6 OR SO3312.  
 OS Shewanella oneidensis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;  
 OC Alteromonadales; Shewanella.  
 OX NCBI\_TaxID=70863;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MR-1;  
 RX MEDLINE=22297686; PubMed=12368813;  
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,  
 RA Read T.D., Eisen J.A., Seehadri R., Ward N., Methe B., Clayton R.A.,  
 RA Meyer R.T., Tappin A., Scott J., Beanan M., Brinkac L., Daugherty S.,  
 RA Debby R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,  
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,  
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,  
 RA Mueller J., Khouri H., Gill J., Utecherback T.R., McDonald L.A.,  
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;  
 RT "genome sequence of the dissimilatory metal ion-reducing bacterium  
 RT Shewanella oneidensis";  
 RL Nat. Biotechnol. 20:1118-1123(2002).  
 DR EMBL; AB015769; AAN56310.1; -;  
 DR TIGR; SO3312; -;  
 KW Complete proteome.  
 SQ SEQUENCE 371 AA; 40603 MW; 024A994C6EC93906 CRC64;

Query Match 84.9%; Score 1584.5; DB 16; Length 371;  
 Best Local Similarity 84.8%; Pred. No. 3.8e-109;  
 Matches 313; Conservative 32; Mismatches 21; Indels 3; Gaps 2;

QY 1 MANOAPIQRRKSTRIVYGNVPIGDAPIAVQSMNTTDTTVEATVNOIKALERGVADIVR 60  
 DB 1 MYNETPIKRPRSTRIVYGNVPIGDAPIAVQSMNTTDTTVEATVNOIKALERGVADIVR 60  
 QY 61 VSVPTMAAFAFKLIKQOVNVPVADIDHFDYRIALKAAYGVDCLRINPNIINBERIRM 120  
 DB 61 VSVPTMAAFAFKLIKQOVNVPVADIDHFDYRIALKAAYGVDCLRINPNIINBERIRM 120  
 QY 121 VVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180  
 DB 121 VVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180  
 QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVKSATGILLSGIGDTLRVSL 240  
 DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVKSATGILLSGIGDTLRVSL 240  
 QY 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMV 300  
 DB 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMV 300

DB 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMV 300  
 QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKAKASQL 359  
 DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKAKASQL 359  
 QY 360 DEARIDVQ 368  
 DB 361 --ANRIQVK 367

## RESULT 3

Q8DIY3 PRELIMINARY; PRT; 366 AA.  
 AC Q8DIY3;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE GCPE protein.  
 GN GCPE  
 OS Wigglesworthia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriales; Wigglesworthia.  
 OX NCBI\_TaxID=164609;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,  
 RA Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse  
 RT flies, Wigglesworthia glosiniidii";  
 RL Nat. Genet. 32:402-407(2002).  
 DR EMBL; AB063522; BAC24719.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 366 AA; 40407 MW; 159E0169B8FC8DED CRC64;

Query Match 65.7%; Score 1226.5; DB 16; Length 366;  
 Best Local Similarity 65.9%; Pred. No. 1.2e-82;  
 Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;

QY 7 IQRRKSTRIVYGNVPIGDAPIAVQSMNTTDTTVEATVNOIKALERGVADIVR 66  
 DB 7 IIRRSKSIYIKVPIGDSPIVQSMNTTCTDINSTISQINKQKAGADIVR 66  
 QY 67 DAAEAFKLIKQOVNVPVADIDHFDYRIALKAAYGVDCLRINPNIINBERIRMVDCAR 126  
 DB 67 EAAEAFKLIKQOVNVPVADIDHFDYRIALKAAYGVDCLRINPNIINBERIRMVDCAR 126  
 QY 127 DKNIPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSVKASDVR 186  
 DB 127 EKLPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSVKASDVR 186  
 QY 187 LAVESYRLAKQIDOPHLGITEAGARGSAVKSATGILLSGIGDTLRVSLADPVE 246  
 DB 187 TCVQSYKTLASKIDQDLHGLITESGMLHGISISGIGILLSGIGDTLRVSLADPVE 246  
 QY 247 EIVYGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMVSIICV 306  
 DB 247 EIVYGFDLKSLRIRSRGINFIACPTCSROEPFVIGTVALLEORLEDDITPMVSIICV 306  
 QY 307 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKAKASQL 360  
 DB 307 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKAKASQL 360

## RESULT 4

Q8G7Y6 PRELIMINARY; PRT; 403 AA.  
 AC Q8G7Y6;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase 1.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:02:08 ; Search time 59 Seconds  
(without alignments)  
1627.044 Million cell updates/sec

Title: US-09-921-992-50  
Perfect score: 166  
Sequence: 1 MHNAPIQRKSTRIVGVN.....RAKASQDEARRIDVQVEK 372

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_23:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_protist:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1631.5	87.4	372	16	Q8DEZ8 vibrio vuln
2	1584.5	84.9	371	16	Q8EC32 shewanella
3	1226.5	65.7	366	16	Q8DIY3 wigglewort
4	814.5	43.6	403	16	Q8G7Y6 bifidobacte
5	753	40.4	393	16	Q8FP82 corynebacte
6	739.5	39.6	362	16	Q8EUI6 mycoplasma
7	601.5	32.2	402	16	Q8DK70 synecococc
8	525.5	28.2	663	16	Q8FIH5 leptospira
9	508	27.2	420	16	Q8FYT2 brucella su
10	471	25.2	740	10	Q8GZK6 lycopersico
11	463	24.8	741	10	Q8RG88 arabidopsis
12	462	24.8	716	10	Q8PFS9 arabidopsis
13	462	24.8	740	10	Q8GZK7 arabidopsis
14	457	24.5	741	10	Q8LPO4 arabidopsis
15	455.5	24.4	746	16	Q8KG23 chlorobium
16	446	23.9	824	5	Q9BJX5 plasmodium

17	446	23.9	824	5	Q8IJH7 plasmodium
18	136	7.3	803	17	Q8R927 thermoaer
19	122.5	6.6	507	17	Q8TZP6 pyrococcus
20	119	6.4	535	16	Q8DJ32 synecococc
21	114.5	6.1	588	2	Q8VSQ9
22	114	6.1	524	17	Q27773 methanobact
23	112.5	6.0	686	17	Q9UXG1 sulfolobus
24	112.5	6.0	723	3	Q9PAD5 emericella
25	111.5	6.0	344	16	Q8EG09 oceanobacti
26	111.5	6.0	578	2	Q68221 anaplasm
27	111.5	6.0	588	2	Q8VSR0 actinobacti
28	111	5.9	575	16	Q8XVK5 raletonia s
29	109	5.8	764	2	Q9RNG5 xanthomonas
30	108.5	5.8	449	16	Q8PZ29 methanomonas
31	108.5	5.8	1000	17	Q8RPA1 lactococcus
32	106.5	5.7	284	16	Q9CHV6 pseudomonas
33	106.5	5.7	798	2	Q9KHS7 pseudomonas
34	106.5	5.7	798	2	Q8KURS pseudomonas
35	106	5.7	748	17	Q57616 methanococc
36	105.5	5.7	578	2	Q93MY9 anaplasm
37	104.5	5.6	638	5	Q9BMQ6 opisthorchi
38	104	5.6	606	16	Q8ITW3 listeria mo
39	104	5.6	1227	16	Q97K41 clostridium
40	103.5	5.5	537	16	Q8FL75 escherichia
41	102.5	5.5	417	3	Q9CA73 emericella
42	102	5.5	546	16	Q8DZ63 streptococ
43	102	5.5	546	16	Q8DX35 streptococ
44	101.5	5.4	564	16	Q44422 heliothis v
45	101.5	5.4	572	10	Q8GZY8 oryza sativ

## ALIGNMENTS

## RESULT 1

ID	Q8DEZ8	PRELIMINARY;	PRT;	372 AA.
AC	Q8DEZ8;			
DT	01-MAR-2003 (TEMBLrel. 23, Created)			
DT	01-MAR-2003 (TEMBLrel. 23, Last sequence update)			
DT	01-MAR-2003 (TEMBLrel. 23, Last annotation update)			
DE	GCPE.			
GN	VV10427.			
OS	Vibrio vulnificus.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;			
OC	Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=672;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRATIN=CMCP6;			
RA	Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,			
RA	Choy H.E.;			
RT	"Complete genome sequence of Vibrio vulnificus CMCP6.";			
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.			
KW	EMBL; AEO16798; AAC08950.1; -			
DR	Complete proteome.			
SQ	SEQUENCE 372 AA; 40573 MW; 6C73C2B8819B0285 CRC64;			
Query Match	87.4%; Score 1631.5; DB 16; Length 372;			
Best Local Similarity	86.7%; Pred. No. 1.3e-112;			
Matches 320; Conservative 29; Mismatches 19; Indels 1; Gaps 1;				
QY	1 MHNAPIQRKSTRIVGVNPIVDGAPPIAVQSMTRTTDVEATVNOIKALEVGGADIVR 60			
DB	1 MHNESEPIIRKSTRIVGVNPIVDGAPPIAVQSMTRTTDVEATVNOIKALEVGGADIVR 60			
QY	61 VSVPTMDAAEAFGLIKQOVNVPVADIHDPYRATLVAEYGVDCILINPNIENERRIM 120			
DB	61 VSVPTMDAAEAFGLIKQOVNVPVADIHDPYRATLVAEYGVDCILINPNIENERRIM 120			
QY	121 VVDCARDKNIPRIGVAGSLEKDLQEKYGEPTPQALLSARHVRHDLRLNDFQKVS 180			
DB	121 VVDCARDKNIPRIGVAGSLEKDLQEKYGEPTPQALLSARHVRHDLRLNDFQKVS 180			

RESULT 15  
ISPG CAUCR STANDARD; PRT; 383 AA.  
AC 09A9W0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
GN ISPG OR CC0851.  
OS Caulobacter crescentus.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;  
OC Caulobacteraceae; Caulobacter.  
OX NCBI\_TaxID=155892;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 19089 / CB15;  
RX MEDLINE=21173698; PubMed=11259647;  
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,  
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,  
RA Potocka I., Nelson R.C., Newton A., Stephens C., Phadke N.D., Ely B.,  
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,  
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,  
RA Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,  
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;  
RT "Complete genome sequence of Caulobacter crescentus";  
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).  
-1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
(ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
(By similarity).  
-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.  
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; AE005761; AKK2836.1; -  
CC DR PIR; H87354; H87354.  
CC DR TIGR; CC0851; -  
CC DR HAMAP; MF\_00159; -; 1.  
CC DR InterPro; IPR004588; ISPG.  
CC DR Pfam; PF04551; GcPB; 1.  
CC DR TIGRFAMs; TIGR00612; ISPG\_gcPB; 1.  
CC KM Isoprene biosynthesis; Complete proteome.  
SQ SEQUENCE 383 AA; 40788 MW; 8468EB400B38194 CRC64;  
Query Match 55.9%; Score 1042.5; DB 1; Length 383;  
Best Local Similarity 57.6%; Pred. No. 9.2e-65;  
Matches 208; Conservative 57; Mismatches 95; Indels 1; Gaps 1;  
QY 7 IORRKTRIVGVNPIIDGAPVAVQSMNTRTDVETVNOIKALEVAGDIYRVSVPTM 66  
DB 14 ITRRQSKRIVGSEVGGDAPISVQSMNTNLTSDAATLTLEQIRLEAGADIVRVSCPDV 73  
QY 67 DAAEAFLLIKQOVNVPVADIHFDYRIALKAEGVVDCLRINPQINERIRRVVDCAR 126  
DB 74 ESTRAFKTIRAEKVPVIVADIIHFHYKGIAPAGAACLRINPQINISPPRVADVIOAKR 133  
QY 127 DKNIPRIGVNAAGLEKDLQEKYGEPTPQALLBSAMRHVDHLDRLNFQFKVSKASDV 186  
DB 134 DHCCSMRIGVNAAGLEKDLQEKYGEPTPQALLBSAMRHVDHLDRLNFQFKVSKASDV 193  
QY 187 LAVESYLLAKQIDQPHLITGEGARSAVKSATIGLLSRTGDTLRVSLADPVE 246  
DB 194 MTAAYYQLEAIDCPHLITGEGATRTGVSATIGIAMLWAGIGDTIRVSLADPVE 253  
QY 247 EIKVGFILKSLIRSRGINFIACPTCSRQEFVIGTVNALQGLEDIITPMVSIIGCV 306

DB 254 EIKVGFILKSLIRSRGINFIACPTCSRQEFVIGTVNALQGLEDIITPMVSIIGCV 313  
QY 307 VNGPGALVSTIGVYGNKSGSL-YEDGYRKRLDNDNDVIDQEAIRAKASQJDEARRI 365  
DB 314 VNGPGALMTDIGFTGGAGAGVYNAKRPDHKQSNQGMIDHIVLVERKKAIEIOAKAO 373  
QY 366 D 366  
DB 374 D 374

Search completed: November 23, 2003, 17:15:28  
Job time : 18 secs

AC P57374; 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR B0287.  
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Buchnera.  
 OK NCBI\_TaxID=118099;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Tokyo 1998;  
 RX MEDLINE=20445173; PubMed=10993077;  
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;  
 RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS.";  
 RL Nature 407:81-86(2000).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AB001118; BAB1297.1; -.  
 DR HAMAP: MF\_00159; -; 1.  
 DR HAMAP: MF\_00159; -; 1.  
 DR InterPro: IPR004588; ISPG.  
 DR Pfam: PF04551; GCPE; 1.  
 DR TIGRFAMs: TIGR00612; ISPG GCPE; 1.  
 KW isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 368 AA; 40869 MW; 88C32F95954C1DE CRC64;  
 Query Match 68.7%; Score 1282; DB 1; Length 368;  
 Best Local Similarity 66.1%; Pred. No. 2.9e-81;  
 Matches 242; Conservative 65; Mismatches .57; Indels 2; Gaps 2;  
 QY 1 MHNQAF-IORRSTRIVGNVPIGDAPIAVOSMTTRTTDEATNOIKALERVAGADIV 59  
 DB 1 MNNKCKIIRNRKSDRIYGVKVAIGNNAPIVSQMTNTRTTINSETINQILEQKVGVDIV 60  
 QY 60 RVSVPTMDAAEAFKLLKQOVNPLVADHFDYRIALKAVEYGVDCRINPNGNREIR 119  
 DB 61 RSTIPMLKAESKEIKQTNVPLADHFDYRLAQAIRKYGDCRINPNGNRRYS 120  
 QY 120 MYVDCARDKINIPRIIGVNGSLEKDOEQEYGEPTPOALLSARHVDHLDRLNFDPKYS 179  
 DB 121 EIIISYAKDENIPRIIGVNGSLEKDLKTKYITPPLALVESAMRHLEFYDALNFDPKYS 180  
 QY 180 VKASDVFLAVESYRLAKQIDQPLHIGITEAGASGAVKSAIGLLSEGIQDTRLVS 239  
 DB 181 VKASDVFLAVESYRLAKQIDQPLHIGITEAGASGAVKSAIGLLSEGIQDTRLVS 240  
 QY 240 LAADPVEETIKVGDIKSLRISRGINFACPTCSQOEFDVIGTVNALQORLEDDITPMD 299  
 DB 241 LAADPVEETIKVGDIKSLRISRGINFACPTCSQOEFDVINTVNALEKNEDEISTPID 300  
 QY 300 VAIIGCVNAGPESALVSTLGTGNGKSGLYEDGV- KDRLDNNMDIDQLEARIKAKSO 358  
 DB 301 VAIIGCVNAGPESALVSTLGTGNGKSGLYEDGV- KDRLDNNMDIDQLEARIKAKSO 360  
 QY 359 LDEARR 364  
 DB 361 LNNKK 366

RESULT 14  
 ID ISPG\_WIGBR STANDARD; PRT; 366 AA.  
 AC Q8D1Y3;  
 DT 15-SEP-2003 (Rel. 42, Created)  
 DT 15-SEP-2003 (Rel. 42, Last sequence update)  
 DT 15-SEP-2003 (Rel. 42, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR WIGBR5730.  
 OS Wigglesworthia glossinidia brevipalpis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Wigglesworthia.  
 OK NCBI\_TaxID=36870;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=22297718; PubMed=12219091;  
 RA Akman L., Yamaehita A., Watanabe H., Oshima K., Shiba T., Hattori M., Aksoy S.;  
 RT "Genome sequence of the endocellular obligate symbiont of tsetse flies, Wigglesworthia glossinidia.";  
 RL Nat. Genet. 32:402-407(2002).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AB063522; BAC24719.1; -.  
 DR HAMAP: MF\_00159; -; 1.  
 DR InterPro: IPR006705; GCPE.  
 DR InterPro: IPR004588; ISPG.  
 DR Pfam: PF04551; GCPE; 1.  
 DR TIGRFAMs: TIGR00612; ISPG GCPE; 1.  
 KW isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 366 AA; 40407 MW; 159B0169B8F0CDED CRC64;  
 Query Match 65.7%; Score 1226.5; DB 1; Length 366;  
 Best Local Similarity 65.9%; Pred. No. 1.9e-77;  
 Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;  
 QY 7 IORRSTRIVGNVPIGDAPIAVOSMTTRTTDEATNOIKALERVAGADIVRSVPTM 66  
 DB 7 IIRRSKSLIYKIVNPVIGGSPISVOSMTNCTTIDNSTISQINKQAKADIVRSIPL 66  
 QY 67 DAEEAFKLLKQOVNPLVADHFDYRIALKAVEYGVDCRINPNGNREIRIMVDCAR 126  
 DB 67 EAAESFKIKRNVSLPIVADHFDYRIALKAVEYGVDCRINPNGNIGKRIISVSTAK 126  
 QY 127 DKNIPRIIGVNGSLEKDOEQEYGEPTPOALLSARHVDHLDRLNFDPKYSVKASDV 186  
 DB 127 EKPLPRIIGVNGSLEKDIENKYGINNPKLPSAMRHVNILEKINFDFKYSVKSSDV 186  
 QY 187 LAVESYRLAKQIDQPLHIGITEAGASGAVKSAIGLLSEGIQDTRLSLAADPVE 246  
 DB 187 TCVQSKYKLAKSLIDQPLHIGITEAGASGAVKSAIGLLSEGIQDTRLSLAADPVE 246  
 QY 247 EIKVGDIKSLRISRGINFACPTCSQOEFDVIGTVNALQORLEDDITPMDVSIIGCV 306  
 DB 247 EYKVGESILRSINIKRGINFACPTCSQOEFDVINNVANVEKRLIEDVTPMNVSVIGCM 306  
 QY 307 VNGPESALVSTLGTGNGKSGLYEDGV- KDRLDNNMDIDQLEARIKAKSO 360  
 DB 307 VNGPESALVSTLGTGNGKSGLYEDGV- KDRLDNNMDIDQLEARIKAKSO 361



RT "whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512 (1995).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: U32721; AAC22026.1; -.  
 CC PIR: H64063; H64063.  
 CC TIGR: H10368; -.  
 CC HAMAP: MF\_00159; -; 1.  
 CC InterPro: IPR004588; Ispg.  
 CC Pfam: PF04551; GCPE; 1.  
 CC TIGRPFAM: TIGR00612; ispg\_gcpe; 1.  
 CC Isoprene biosynthesis; Complete proteome.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 368 AA; 40116 MW; DC99EAL5A50ED01B CRC64;

Query Match 85.9%; Score 1603.5; DB 1; Length 368;  
 Best Local Similarity 88.2%; Pred. No. 2.1e-103;  
 Matches 320; Conservative 20; Mismatches 22; Indels 1; Gaps 1;

QY 4 QAPIORKSTRIVGNVPIGDCAPPIAVOSMTNRTTDEATVNOIKALERVGADIVRSV 63  
 DB 5 OPTIKRRESTKIYGVNVPPIGDCAPPIAVOSMTNRTTDEATVNOIKALERVGADIVRSV 64  
 QY 64 PTMDAAEAKRLIKQOVNPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMYD 123  
 DB 65 PTMDAAEAKRLIKQOVNPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMYD 124  
 QY 124 CARDKPIRIGVNGASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKVSVKAS 183  
 DB 125 CARDKPIRIGVNGASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKVSVKAS 184  
 QY 184 DVFLLAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAAD 243  
 DB 185 DVFLLAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAAD 244  
 QY 244 PVEEIVGVFDILKSLRISRGINFIACPTCSROEPFVIGTVNALRQRLIEDITPMDSII 303  
 DB 245 PVEEIVGVFDILKSLRISRGINFIACPTCSROEPFVIGTVNALRQRLIEDITPMDSII 304  
 QY 304 GGVNGPGEALVSTLGVNGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 362  
 DB 305 GGVNGPGEALVSDLGVTGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 364  
 QY 362 GGVNGPGEALVSDLGVTGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 364  
 DB 363 GGVNGPGEALVSDLGVTGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 364  
 QY 363 GGVNGPGEALVSDLGVTGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 364  
 DB 363 GGVNGPGEALVSDLGVTGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 364

RESULT 10  
 ISPG\_PASMU STANDARD; PRT; 367 AA.  
 ID ISPG\_PASMU  
 AC P57987;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR PM2010.  
 OS Pasteurella multocida.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
 OC Pasteurellaceae; Pasteurella.  
 NCBI\_TaxID=747;

RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=PM70;  
 RX MEDLINE=21145866; PubMed=11248100;  
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
 RT "Complete genomic sequence of Pasteurella multocida PM70.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 CC (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
 CC (By similarity).  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AE006237; AAK0494.1; -.  
 CC HAMAP: MF\_00159; -; 1.  
 CC InterPro: IPR004588; Ispg.  
 CC Pfam: PF04551; GCPE; 1.  
 CC TIGRPFAM: TIGR00612; ispg\_gcpe; 1.  
 CC Isoprene biosynthesis; Complete proteome.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 367 AA; 39969 MW; 1918C5702642AE8 CRC64;

Query Match 85.7%; Score 1599.5; DB 1; Length 367;  
 Best Local Similarity 88.9%; Pred. No. 4e-103;  
 Matches 320; Conservative 18; Mismatches 21; Indels 1; Gaps 1;

QY 7 IQRRKSTRIVGNVPIGDCAPPIAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTM 66  
 DB 7 IQRRKSTRIVGNVPIGDCAPPIAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTM 66  
 QY 67 DAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMYD 126  
 DB 67 DAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCIRINPGNIGNEERIRMYD 126  
 QY 127 DKNIPRIGVNGASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKVSVKASDVF 186  
 DB 127 DKNIPRIGVNGASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKVSVKASDVF 186  
 QY 187 LAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAADPVE 246  
 DB 187 LAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAADPVE 246  
 QY 247 EIKVGFDIKSLRISRGINFIACPTCSROEPFVIGTVNALRQRLIEDITPMDSII 306  
 DB 247 EIKVGFDIKSLRISRGINFIACPTCSROEPFVIGTVNALRQRLIEDITPMDSII 306  
 QY 307 VNGPGEALVSTLGVNGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 365  
 DB 307 VNGPGEALVSDLGVTGKKSGLYEDGVR-KDRLNNDMIDQLEAIRAKASQLDEAR 366

RESULT 11  
 ISPG\_PSEAE STANDARD; PRT; 371 AA.  
 ID ISPG\_PSEAE  
 AC Q9HXJ4;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR PA3803.  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NCBI\_TaxID=287;  
 RN [1]  
 RP SEQUENCE FROM N.A.



KW Isoprene biosynthesis; Complete proteome.  
SQ SEQUENCE 375 AA; 40797 MW; 860971872B1C0536 CRC64;

Query Match 87.0%; Score 1622.5; DB 1; Length 375;

Best Local Similarity 86.8%; Pred. No. 1.1e-104; Mismatches 29; Indels 1; Gaps 1;

Matches 323; Conservative 19; Mismatches 29; Indels 1; Gaps 1;

QY 1 MHNQAFIQRRKSTRIVGVNPIGDGAPIAVQSMNTTRTTDVATVQIKALERVGADIVR 60  
DB 1 MHNQAFIIIRKSTRIVGVNPIGDGAPIAVQSMNTTRTTDVATVQIKALERVGADIVR 60  
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 120  
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 120  
QY 121 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 180  
DB 121 VVACARDYNIPIRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 180  
QY 181 KASDVFLAVESYRLAKQIDQPLHLGITEAGARSAVSAIGLLSEGIQDTRVSL 240  
DB 181 KASDVFLAVNSYRLAKQIDQPLHLGITEAGARSAVSAIGLLSEGIQDTRVSL 240  
QY 241 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNLEQRLIEDITPM 300  
DB 241 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNLEQRLIEDITPM 300  
QY 301 SIIGCVNNGPGEALVSTLGTGANKKSGLYEGVGR-KDRLDNNMDLDLEARRARASQ 359  
DB 301 SIIGCVNNGPGEALVSTLGTGANKKSGLYEGVGR-KDRLDNNMDLDLEARRARASQ 359  
QY 360 DEARRIDVOQVE 371  
DB 361 DANNRIVINQLD 372

## RESULT 8

ISPG\_VIBCH STANDARD; PRT; 376 AA.

AC 09KTYL, 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
GN ISPG OR VC0759.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=El Tor N16961 / Serotype O1;  
RX MEDLINE=20406833; PubMed=10952301;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,  
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,  
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,  
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
RA McDonald L., Uetshack T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae";  
RT Nature 406:477-483(2000).  
RL - FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
CC (ME-2,4CpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate  
CC (BY similarity).  
CC -!- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.  
CC -!- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, AF004161; AAF93924.1; --  
DR PIR; F82283; F82283.  
DR TIGR; VC0759; --  
DR HAMAB; MF\_00159; --  
DR InterPro; IPR004588; ISPG.  
DR Pfam; PF04551; Gcpe; 1.  
DR TIGRPFAMs; TIGR00612; ISPG\_gcpe; 1.  
KW Isoprene biosynthesis; Complete proteome.  
SQ SEQUENCE 376 AA; 40862 MW; 34E4144CB87ADAAA CRC64;

Query Match 86.9%; Score 1621.5; DB 1; Length 376;  
Best Local Similarity 85.3%; Pred. No. 1.3e-104;  
Matches 318; Conservative 33; Mismatches 21; Indels 1; Gaps 1;

QY 1 MHNQAFIQRRKSTRIVGVNPIGDGAPIAVQSMNTTRTTDVATVQIKALERVGADIVR 60  
DB 3 MHNQAFIIIRKSTRIVGVNPIGDGAPIAVQSMNTTRTTDVATVQIKALERVGADIVR 62  
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 120  
DB 63 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 122  
QY 121 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 180  
DB 123 VVDCARDKNIPIRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 182  
QY 181 KASDVFLAVESYRLAKQIDQPLHLGITEAGARSAVSAIGLLSEGIQDTRVSL 240  
DB 183 KASDVFLAVDSYRLAKQIDQPLHLGITEAGARSAVSAIGLLSEGIQDTRVSL 242  
QY 241 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNLEQRLIEDITPM 300  
DB 243 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNLEQRLIEDITPM 302  
QY 301 SIIGCVNNGPGEALVSTLGTGANKKSGLYEGVGR-KDRLDNNMDLDLEARRARASQ 359  
DB 303 SIIGCVNNGPGEALVSTLGTGANKKSGLYEGVGR-KDRLDNNMDLDLEARRARASQ 362  
QY 360 DEARRIDVOQVE 372  
DB 363 DEKRRIDVQVQ 375

## RESULT 9

ISPG\_HAEIN STANDARD; PRT; 368 AA.

AC P44667, 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
GN ISPG OR GCPE OR HI0368.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=127;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Rd / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uetshack T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;

```

QY 301 SIIGCVNPGGALVSTLGTGNNKSGLYEDGVR-KDRLDNNMDIDLEARIKASQUL 359
DB 301 SIIGCVNPGGAEVSHLGLAGSNKKSAYEDQKRFEDNDLYNQLEAKIRAKAAM 360
QY 360 DEARRIDVO 368
DB 361 DESNRIDIK 369

RESULT 6
ISPG_PROST STANDARD; PRT; 365 AA.
ID ISPG_PROST
AC P72241;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR AAC.
OS Providencia stuartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxId=588;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=PR50;
RX MEDLINE=97234638; PubMed=9079912;
RA Rather P.N., Solinsky K.A., Paradise M.R., Parojic M.M.;
RT "aacC, an essential gene involved in density-dependent regulation of
RT the 2',N-acetyltransferase in Providencia stuartii."
RL J. Bacteriol. 179:2267-2273(1997).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cnp) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity). Involved in density-dependent regulation of 2'-N-
CC acetyltransferase.
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67933; AAB51469.1; -
CC HAMAP; MF_00159; -; 1.
CC DR InterPro; IPR004588; ISPG.
CC DR Pfam; PF04551; GCPE; 1.
CC DR TIGRFAMs; TIGR00612; ispg_gcpe; 1.
CC KM Isoprene biosynthesis.
CC SEQUENCE 365 AA; 39840 MW; 4523147980480ECC CRC64;

Query Match 87.0%; Score 1622.5; DB 1; Length 365;
Best Local Similarity 88.4%; Pred. No. 1e-104;
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;

```

```

QY 241 AADPVEIKVGPDIKSLIRSRGINFLACPTRQDFVIGTVNALEQRLEDIITPMY 300
DB 241 AADPVEIVGVGDILKSLIRSRGINFLACPTRQDFVIGTVNALEQRLEDIITPMY 300
QY 301 SIIGCVNPGGALVSTLGTGNNKSGLYEDGVR-KDRLDNNMDIDLEARIKASQUL 359
DB 301 SIIGCVNPGGAEVSHLGLAGSNKKSAYEDQKRFEDNDLYNQLEAKIRAKAAM 360
QY 360 DE 361
DB 361 DE 362

RESULT 7
ISPG_YERPE STANDARD; PRT; 375 AA.
ID ISPG_YERPE
AC P58672;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR GCPE OR IPO2879 OR Y1353.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxId=632;
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parthill J., Wren B.W., Thomson N.R.,
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Meule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN (2)
RN SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=2213763; PubMed=12142430;
RA Deng M., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.T., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cnp) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ41454; CAC92130.1; -
CC EMBL; AB013738; AAM84926.1; -
CC PIR; AG0350; AG0350.
CC DR HAMAP; MF_00159; -; 1.
CC DR InterPro; IPR006705; GCPE.
CC DR InterPro; IPR004588; ISPG.
CC DR Pfam; PF04551; GCPE; 1.
CC DR TIGRFAMs; TIGR00612; ispg_gcpe; 1.

```

```

Db      301 SIIGCVNNGPGEALVSTLGVGTGANKKSGLYEDGVKRDLDNDMMIAQLSRIKAVXSQLD 360
Qy      361 EARRIDVQVEK 372
Db      361 EARRIDVQVEK 372

RESULT 4
ISPG_VIBPA STANDARD; PRT; 372 AA.
ID      ISPG_VIBPA 087516;
AC      15-SEP-2003 (Rel. 42, Created)
DT      15-SEP-2003 (Rel. 42, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
GN      1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
OS      ISPG OR VP0608.
OC      Vibrrio parahaemolyticus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrrio.
NCBI_TaxID=670;

RP      SEQUENCE FROM N.A.
RC      STRAIN=RIMD 2210633 / Serotype O3:K6;
RX      MEDLINE=22508454; Pubmed=12620739;
RA      Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA      Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA      Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT      "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT      distinct from that of V. cholerae.";
RL      Lancet 361:743-749(2003).
CC      -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC      (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC      (By similarity).
CC      -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC      -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AP05075; BACS8871.1; -.
DR      HAMAP; MF_00159; -; 1.
KM      Isoprene biosynthesis; Complete proteome.
SQ      SEQUENCE 372 AA; 40688 MW; EAAFAAALJA764DDB8 CRC64;

Query Match      88.2%; Score 1645.5; DB 1; Length 372;
Best Local Similarity 87.5%; Pred. No. 2.8e-106;
Matches 323; Conservative 27; Mismatches 18; Indels 1; Gaps 1;

Qy      1 MHNQAPIORRKRTRIVGVNPIGGAPIAVQSMNTTTRTVEATVNVQIKALERGADIVR 60
Db      1 M0HESPPIRRKSRTRIVGDVPIGGAPIAVQSMNTTTRTVEATVNVQIKALERGADIVR 60
Qy      61 VSVPTMDAAEAFKLIKQOVNVPVADIHFDYRIALKVAEYGVDCLRINPNIIGNEERIRM 120
Db      61 VSVPTMDAAEAFKLIKQOVNVPVADIHFDYRIALKVAEYGVDCLRINPNIIGNEERIRS 120
Qy      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Db      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Qy      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Db      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Qy      181 KASDVFLAVESYRLAKIDQPLHLGITEAGARGAVKSAIGGLLSGIGDTLRVSL 240
Db      181 KASDVFLAVESYRLAKIDQPLHLGITEAGARGAVKSAIGGLLSGIGDTLRVSL 240
Qy      241 AADPVEIKVGFILKSLIRSRGINFIACPTCSROEFDVIGVNALEORLEDITITPMDV 300
Db      241 AADPVEIKVGFILKSLIRSRGINFIACPTCSROEFDVIGVNALEORLEDITITPMDV 300

```

```

Qy      301 SIIGCVNNGPGEALVSTLGVGTGANKKSGLYEDGVKRDLDNDMMIAQLSRIKAVXSQLD 359
Db      301 SIIGCVNNGPGEALVSTLGVGTGANKKSGLYEDGVKRDLDNDMMIAQLSRIKAVXSQLD 360
Qy      360 DEARRIDVQ 368
Db      361 DSENRIEIK 369

RESULT 5
ISPG_VIBVU STANDARD; PRT; 372 AA.
ID      ISPG_VIBVU 08828;
AC      15-SEP-2003 (Rel. 42, Created)
DT      15-SEP-2003 (Rel. 42, Last sequence update)
DE      15-SEP-2003 (Rel. 42, Last annotation update)
GN      1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
OS      ISPG OR VV10427.
OC      Vibrio vulnificus.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC      Vibrionaceae; Vibrrio.
NCBI_TaxID=672;

RP      SEQUENCE FROM N.A.
RC      STRAIN=CMCPE;
RX      Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA      Choy H.E.;
RT      "Complete genome sequence of Vibrio vulnificus CMCPE.";
RT      Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RL      -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC      (ME-2,4cPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC      (By similarity).
CC      -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC      -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE016798; AAO08950.1; -.
DR      HAMAP; MF_00159; -; 1.
DR      Pfam; PF04551; GcPE; 1.
DR      TIGRPFAMs; TIGR00612; ispg gcPE; 1.
KM      Isoprene biosynthesis; Complete proteome.
SQ      SEQUENCE 372 AA; 40573 MW; 6C73C2B8819B0285 CRC64;

Query Match      87.4%; Score 1631.5; DB 1; Length 372;
Best Local Similarity 86.7%; Pred. No. 2.6e-105;
Matches 320; Conservative 29; Mismatches 19; Indels 1; Gaps 1;

Qy      1 MHNQAPIORRKRTRIVGVNPIGGAPIAVQSMNTTTRTVEATVNVQIKALERGADIVR 60
Db      1 M0HESPPIRRKSRTRIVGDVPIGGAPIAVQSMNTTTRTVEATVNVQIKALERGADIVR 60
Qy      61 VSVPTMDAAEAFKLIKQOVNVPVADIHFDYRIALKVAEYGVDCLRINPNIIGNEERIRM 120
Db      61 VSVPTMDAAEAFKLIKQOVNVPVADIHFDYRIALKVAEYGVDCLRINPNIIGNEERIRS 120
Qy      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Db      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Qy      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Db      121 VVDCARDKNIPIRIGVNGAGSLKDLQKYGEPPEALVESAMRHVDLDRINFDPQFVSV 180
Qy      181 KASDVFLAVESYRLAKIDQPLHLGITEAGARGAVKSAIGGLLSGIGDTLRVSL 240
Db      181 KASDVFLAVESYRLAKIDQPLHLGITEAGARGAVKSAIGGLLSGIGDTLRVSL 240
Qy      241 AADPVEIKVGFILKSLIRSRGINFIACPTCSROEFDVIGVNALEORLEDITITPMDV 300
Db      241 AADPVEIKVGFILKSLIRSRGINFIACPTCSROEFDVIGVNALEORLEDITITPMDV 300

```

```

RT      LT2." ;
RL      Nature 413:852-856(2001).
CC      -l- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC      (ME-2,4cpe) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC      (By similarity).
CC      -l- PATHWAY: Nucleoside terepoid biosynthesis pathway; sixth step.
CC      -l- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AE008814; AAL21417.1; -.
DR      StyGene; SG7777; ISPG.
DR      HAMAP; MF_00159; -; 1.
DR      InterPro; IPR006705; GcPE.
DR      InterPro; IPR004588; ISPG.
DR      Pfam; PF04551; GcPE; 1.
DR      TIGRFAMs; TIGR00612; ISPG_GcPE; 1.
DR      KEGG; isoprene_biosynthesis; Complete proteome.
SQ      SEQUENCE 372 AA; 40625 MW; 0486FD3BECA2AC9F CRC64;

Query Match      98.6%; Score 1840; DB 1; Length 372;
Best Local Similarity 98.7%; Pred. No. 1,1e-119;
Matches 367; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 MHNQAPIORRSTIYGVNVPVLAADHFDYRIALKVAEYGVDCIRINPGNGEERIRM 60
DB      1 MHNQAPIORRSTRIYGVNVPVIGGAPVAVGSMNTRRTDVEATVNOIKALERGADIVR 60
QY      61 VSVPTMAAEEAFKIKOOVNPVLAADHFDYRIALKVAEYGVDCIRINPGNGEERIRM 120
DB      61 VSVPTMAAEEFKLIKOOVNPVLAADHFDYRIALKVAEYGVDCIRINPGNGEERIRM 120
QY      121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTQALLSARHVDLDRINPQFQSV 180
DB      121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTQALLSARHVDLDRINPQFQSV 180
QY      181 KASVPLAVESYRLAAQIDQPHLGITTEAGASGAVKSAIGLLISGIGTLRVSL 240
DB      181 KASVPLAVESYRLAAQIDQPHLGITTEAGASGAVKSAIGLLISGIGTLRVSL 240
QY      241 AADVEEIKYGFDFLKSRLRSRGINFACPTSGROEFDVGTNALRQRLIEDITPMDV 300
DB      241 AADVEEIKYGFDFLKSRLRSRGINFACPTSGROEFDVGTNALRQRLIEDITPMDV 300
QY      301 SIICVNVNGEALVSTLGVTKGNGKSGLYEDVGRKQRLDNDNDVIDLEAIRAKASQLD 360
DB      301 SIICVNVNGEALVSTLGVTKGNGKSGLYEDVGRKQRLDNDNDVIDLEAIRAKASQLD 360
QY      361 EARRIDVQVEK 372
DB      361 EARRIDVQVEK 372
DB      361 EARRIDVQVEK 372

RESULT 3
ISPG_SALT1
ID      ISPG_SALT1 STANDARD; PRT; 372 AA.
AC      P58670;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN      ISPG OR GcPE OR STY2768 OR T0333.
OS      Salmonella typhi.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Salmonella.
NCBI_TaxId=601;
LN      [1]

```

RC	SEQUENCE FROM N.A.
RA	MEDLINE=21534947; PubMed=11677608;
RX	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA	Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA	Krogh A., Rutherford T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA	Quail M., Ratheford K., Simmonds M., Skelton J., Stevens K.,
RA	Whitehead S., Barrrell B.G.;
RT	"Complete genome sequence of a multiple drug resistant Salmonella
RL	enterica serovar Typhi CT18.",
RU	Nature 413:848-852(2001).
RU	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=Ty2 / ATCC 700931;
RX	MEDLINE=22531367; PubMed=12664504;
RA	Deng W., Lion S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,
RA	Butland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT	"Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RL	and CT18.";
CC	J. Bacteriol. 185:2330-2337(2003).
CC	-1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC	(ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC	(by similarity).
CC	-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC	-1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/">http://www.isb.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL; AL627275; CAD02726.1; -;
DR	EMBL; AB016835; AA068056.1; -;
DR	HAMAP; MF_00159; -; 1.
DR	InterPro; IPR006705; GcPE.
DR	InterPro; IPR004588; Ispg.
DR	Pfam; PF04551; Gcpe; 1.
DR	TIGRFAMS; TIGR00612; ispg_gcpe; 1.
DR	Isoprene biosynthesis; Complete proteome.
KM	Sequence 372 AA; 40626 MW; 3E3DSB13BAHA7916 CRC64;
Query Match	98.1%; Score 1831; DB 1; Length 372;
Best Local Similarity	98.1%; Pstd. No. 4.8e-119;
Matches 365; Conservative	4; Mismatches 3; Indels 0; Gaps 0;
OY	1 MHNOAPIORCKSRIRIVGNVPIGDGAPIAVQSMNTNRTTVETVAOVOKALERVGADIVR 60
DB	1 MHNOAPIORCKSRIRIVGNVPIGDGAPIAVQSMNTNRTTVETVAOVOKALERVGADIVR 60
OY	61 VSVPTDAEAFAFLIKQQVNVPLVADIHPDYRALKVAEGVDCLRNPNIGNEEIRM 120
DB	61 VSVPTDAEAFAFLIKQQVNVPLVADIHPDYRALKVAEGVDCLRNPNIGNEEIRM 120
OY	121 VVDCARDKNIPIRIGVNASLEKDLOEKXGEPTPQALLSESMRHVDHLDFLNFDQFVSV 180
DB	121 VVDCARDKNIPIRIGVNASLEKDLOEKXGEPTPQALLSESMRHVDHLDFLNFDQFVSV 180
OY	181 KASDVLAVESYLLAKQIDQPLHITTEAGARGSGAVSAIGLGILLSGICGTDLRVSL 240
DB	181 KASDVLAVESYLLAKQIDQPLHITTEAGARGSGAVSAIGLGILLSGIDTLRVSL 240
OY	241 AADPVEIIVGFPIILSKLRISRGINFACPTCSROEFVIGTVNALLEGLEDIIITPMDV 300
DB	241 AADPVEIIVGFPIILSKLRISRGINFACPTCSROEFVIGTVNALLEGLEDIIITPMDV 300
OY	301 SIIGCVNGPGEALVSTLTGVGNKKSGLYEDVGVRDRLDNNDMDIOLAEIRAKASOLD 360
DB	301 SIIGCVNGPGEALVSTLTGVGNKKSGLYEDVGVRDRLDNNDMDIOLAEIRAKASOLD 360

RP SEQUENCE FROM N.A.  
 RC STRAIN-06.H / CFT073 / ATCC 700928;  
 RA MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G., III, Redford P., Resch P.,  
 RA Maso D., Buckles E.L., Liu S.-R., Boutin J., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 of uropathogenic *Escherichia coli*.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
 RN [6].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157.H7 / BDL933 / ATCC 700927;  
 RA MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grobeck E.J., Davis N.W., Lim A., Dimailanta E.T., Potamianos K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";  
 RL Nature 409:529-533(2001).  
 RN [7].  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-0157.H7 / RIMD 0509952;  
 RA MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*  
 O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 RN [8].  
 RP PRELIMINARY SEQUENCE OF 343-372 FROM N.A.  
 RA MEDLINE=85261421; PubMed=2991272;  
 RA Freedman R., Gibson B., Donovan D., Blemann K., Eisenbeis S.J.,  
 RA Parker J., Schimmel P.;  
 RT "Primary structure of histidine-tRNA synthetase and characterization  
 of his transcripts.";  
 RL J. Biol. Chem. 260:10063-10068(1985).  
 RN [9].  
 RP PATHWAY.  
 RC STRAIN-K12 / MC4100;  
 RA MEDLINE=21099853; PubMed=11163766;  
 RA Campos N., Rodriguez-Concepcion M., Seemann M., Rohner M., Boronat A.;  
 RT "Identification of gcpe as a novel gene of the 2-C-methyl-D-erythritol  
 4-phosphate pathway for isoprenoid biosynthesis in *Escherichia coli*.";  
 RL FEBS Lett. 488:170-173(2001).  
 RN [10].  
 RP PATHWAY.  
 RA MEDLINE=21172855; PubMed=11274098;  
 RA Altincicek B., Kollas A.-K., Sanderbrand S., Wiesner J., Hintz M.,  
 RA Beck E., Jomaa H.;  
 RT "Gcpe is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway  
 of isoprenoid biosynthesis in *Escherichia coli*.";  
 RL J. Bacteriol. 183:2411-2416(2001).  
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate  
 (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.  
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.  
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; X64451; CAA45783.1; -  
 DR EMBL; AY033515; AAK53460.1; -  
 DR EMBL; AE000338; AAC75568.1; -

DR EMBL; D90880; BAA16402.1; -  
 DR EMBL; D90881; BAA20919.1; -  
 DR EMBL; AE016764; FAN81487.1; -  
 DR EMBL; AE005481; AAG57625.1; -  
 DR EMBL; AP002561; BAB36800.1; -  
 DR EMBL; M1843; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; A91051; A91051.  
 DR PIR; E85895; E85895.  
 DR PIR; S23058; S23058.  
 DR EcoGene; EG10370; -spg.  
 DR HAMAP; MF\_00159; -; 1.  
 DR InterPro; IPR004588; Ispg.  
 DR Pfam; PF04551; GCPE; 1.  
 DR TIGRfams; TIGR00612; ispg\_gcpe; 1.  
 KW Isoprene biosynthesis; Complete proteome.  
 SQ SEQUENCE 372 AA; 40683 MW; 3666F83D5CD4P9F CRC64;  
 Query Match 100.0%; Score 1866; DB 1; Length 372;  
 Best local Similarity 100.0%; Pred. No. 1.9e-121;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MNNQAPIQRKSTRIYGVNPIGDGAPLAVQSMNTTTRTDTVEATVNOIALERVGADIYR 60  
 DB 1 MNNQAPIQRKSTRIYGVNPIGDGAPLAVQSMNTTTRTDTVEATVNOIALERVGADIYR 60  
 QY 61 VSPPTMDAAEARKLTKQVNVPLVADHPDYRIALKVAYGVDCIRINGNIGNERIRM 120  
 DB 61 VSPPTMDAAEARKLTKQVNVPLVADHPDYRIALKVAYGVDCIRINGNIGNERIRM 120  
 QY 121 VVDCARDKNIPIRIGVNASLEKDLQEKGEPTPOLLESARHVDLRLNFDQPKVSV 180  
 DB 121 VVDCARDKNIPIRIGVNASLEKDLQEKGEPTPOLLESARHVDLRLNFDQPKVSV 180  
 QY 181 KASDVFLLAVESYRLAKQIDQPLHLGITBAGGARSQVSAIGLLSEGGITLRSVL 240  
 DB 181 KASDVFLLAVESYRLAKQIDQPLHLGITBAGGARSQVSAIGLLSEGGITLRSVL 240  
 QY 241 AADPVEIKVGRFDILKSLRIRSGINFIACPTCSRQEPFVIGVNLLEQRLDITPMQV 300  
 DB 241 AADPVEIKVGRFDILKSLRIRSGINFIACPTCSRQEPFVIGVNLLEQRLDITPMQV 300  
 QY 301 SIICGVNPGGALVSTLGVGTGANKSGLYEGVGRKDRLDNNDMDIQLEARIKASQLD 360  
 DB 301 SIICGVNPGGALVSTLGVGTGANKSGLYEGVGRKDRLDNNDMDIQLEARIKASQLD 360  
 QY 361 EARRIDVQVEK 372  
 DB 361 EARRIDVQVEK 372  
 RESULT 2  
 ID ISPG\_SALTY STANDARD; PRT; 372 AA.  
 AC PS6671;  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.  
 GN ISPG OR GCPE OR STM2523.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Salmonella.  
 OX NCBI\_TaxId=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;  
 RA MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 15:55:37 ; Search time 17 Seconds

(without alignments)  
1029.055 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866  
Sequence: 1 MHNOAPIQRKSTRIVGVN.....RAKASQDARIDVQOVER 372

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_41.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1866	100.0	372	1 ISPG_ECOLI P27433 escherichia
2	1840	98.6	372	1 ISPG_SALTY P58671 salmonella
3	1831	98.1	372	1 ISPG_SALTY P58670 salmonella
4	1645.5	88.2	372	1 ISPG_VIBPA O87816 vibrio para
5	1631.5	87.4	372	1 ISPG_VIBV O8de8 vibrio vuln
6	1622.5	87.0	365	1 ISPG_PROST P72241 providencia
7	1622.5	87.0	375	1 ISPG_YERPE P58672 yersinia pe
8	1621.5	86.9	376	1 ISPG_VIBCH O9KX1 vibrio chol
9	1603.5	85.9	368	1 ISPG_HABIN P44667 haemophilus
10	1599.5	85.7	367	1 ISPG_PASMU P57987 pasteurella
11	1372.5	73.6	371	1 ISPG_PSEAE O8XJ4 pseudomonas
12	1285.5	68.9	367	1 ISPG_BUCAP O8K94 buchnera ap
13	1282	68.7	368	1 ISPG_BUCAI P57334 buchnera ap
14	1226.5	65.7	366	1 ISPG_WIGBR O8d1y3 wigleswort
15	1042.5	55.9	383	1 ISPG_CAUCC O9a990 caulobacter
16	906	48.6	349	1 ISPG_CLOPE P58667 clostridium
17	902.5	48.4	355	1 ISPG_THETN O8a30 thermoaer
18	887	47.5	357	1 ISPG_CLOAB O9J16 clostridium
19	882.5	47.3	367	1 ISPG_BACHD O9K18 bacillus ha
20	869.5	46.6	354	1 ISPG_FUSNO O8G40 fusobacteri
21	868	46.5	368	1 ISPG_LISMO P58668 listeria mo
22	859.5	46.1	377	1 ISPG_BACSU P54482 bacillus su
23	838	44.4	357	1 ISPG_AQUAE O67496 aquifex aeo
24	815.5	43.7	384	1 ISG1_STRCO O9X7W2 streptomyce
25	815	43.6	385	1 ISG2_STRCO O9X7Y9 streptomyce
26	794	42.6	359	1 ISPG_HELPU O25332 helicobacte
27	790	42.3	359	1 ISPG_HELPU O92110 helicobacte
28	761	40.8	378	1 ISPG_CORGL O8n12 corynebacte
29	751.5	40.3	387	1 ISPG_MYCTU O3350 mycobacteri
30	712.5	38.2	392	1 ISPG_MYCTU O9CDB5 mycobacteri
31	706	37.0	344	1 ISPG_THBMA O9WZ3 thermotoga
32	690	37.0	357	1 ISPG_CAMJB O9Pm1 campylobact
33	618.5	33.1	403	1 ISPG_SYNY3 P73672 synchocyst

RESULT 1  
ID ISPG\_ECOLI STANDARD; PRT; 372 AA.  
AC P27433; P76984; P76985;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (Gcpe protein)  
GN (Protein E).  
DE ISPG OR GCPE OR B2515 OR C3037 OR Z3778 OR ECS3377.  
OS Escherichia coli,  
OS Escherichia coli, O6, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562, 217992, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=92394488; PubMed=1521767;  
RA Baker J., Franklin D.B., Parker J.;  
RT "Sequence and characterization of the gcpe gene of Escherichia coli.";  
RL FEMS Microbiol. Lett. 73:1175-180(1992).  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC MEDLINE=97426617; PubMed=9278503;  
RX Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1233-1238(1997).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12;  
RX MEDLINE=97349980; PubMed=9205837;  
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,  
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,  
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,  
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,  
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,  
RA Yamagata S., Horuchi T.;  
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli  
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and  
RT DNA Res. 4:91-113(1997).  
RN [5]

THIS PAGE BLANK (USPTO)











```
QY 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTDTVEATVNOIKALERVGADIYR 60
DB 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTDTVEATVNOIKALERVGADIYR 60
QY 61 VSPPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPQINIGNEERIRM 120
DB 61 VSPPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPQINIGNEERIRM 120
QY 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDOFKVSV 180
DB 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVSAIGLILSEGIGDTLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVSAIGLILSEGIGDTLRVSL 240
QY 241 AADPVEEIKVGFPIILKSIRSRGINPIACPTCSRQEPDVIGTVNALRQLEDITITPMV 300
DB 241 AADPVEEIKVGFPIILKSIRSRGINPIACPTCSRQEPDVIGTVNALRQLEDITITPMV 300
QY 301 SIIGCVNNGPGEALVSTLGVGNGKSGLYEDGVKRDLDNDMDIOLLEARIKASQOLD 360
DB 301 SIIGCVNNGPGEALVSTLGVGNGKSGLYEDGVKRDLDNDMDIOLLEARIKASQOLD 360
QY 361 EARRIDVOQVEK 372
DB 361 EARRIDVOQVEK 372
```

RESULT 5  
AG0350  
probable acetyltransferase aarc [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis  
C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #ext\_change 09-Nov-2001  
C/Accession: AG0350  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.;  
deno-Tarata, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, L.  
Nature 413, 523-527, 2001  
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.  
A/Reference number: AB0001; MUID:21470413; PMID:11586360  
A/Accession: AG0350  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-375 <KUR>  
A/Cross-references: GB:AL590842; PIDN:CA092130.1; PID:g15980846; GSPDB:GN00175  
C/Genetics:  
A/Gene: aarc  
C/Superfamily: gcpb protein

Query Match 87.0%; Score 1621.5; DB 2; Length 375;  
Best Local Similarity 86.8%; Pred. No. 9.1e-105;  
Matches 333; Conservative 19; Mismatches 29; Indels 1; Gaps 1;

```
QY 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTDTVEATVNOIKALERVGADIYR 60
DB 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTDTVEATVNOIKALERVGADIYR 60
QY 61 VSPPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPQINIGNEERIRM 120
DB 61 VSPPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPQINIGNEERIRM 120
QY 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDOFKVSV 180
DB 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDOFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVSAIGLILSEGIGDTLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVSAIGLILSEGIGDTLRVSL 240
QY 241 AADPVEEIKVGFPIILKSIRSRGINPIACPTCSRQEPDVIGTVNALRQLEDITITPMV 300
DB 241 AADPVEEIKVGFPIILKSIRSRGINPIACPTCSRQEPDVIGTVNALRQLEDITITPMV 300
```

```
QY 301 SIIGCVNNGPGEALVSTLGVGNGKSGLYEDGVKRDLDNDMDIOLLEARIKASQOLD 359
DB 301 SIIGCVNNGPGEALVSTLGVGNGKSGLYEDGVKRDLDNDMDIOLLEARIKASQOLD 359
QY 360 DEARRIDVOQVE 371
DB 360 DEARRIDVOQVE 371
```

RESULT 6

F82283  
gcpb protein VC0759 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C/Species: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #ext\_change 02-Feb-2001  
C/Accession: F82283  
R:Reiherberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: AF0035; MUID:20406833; PMID:10952301  
A/Accession: F82283  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-376 <HEI>  
A/Cross-references: GB:AE004161; GB:AE003952; NID:g9655200; PIDN:AAF93924.1; GSPDB:GN001;  
A/Experimental source: serogroup O1; strain N16961; biotype El Tor  
C/Genetics:  
A/Gene: VC0759  
A/Map position: 1  
C/Superfamily: gcpb protein

Query Match 86.9%; Score 1621.5; DB 2; Length 376;  
Best Local Similarity 85.3%; Pred. No. 1.1e-104;  
Matches 318; Conservative 33; Mismatches 21; Indels 1; Gaps 1;

```
QY 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTDTVEATVNOIKALERVGADIYR 60
DB 3 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTDTVEATVNOIKALERVGADIYR 62
QY 61 VSPPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPQINIGNEERIRM 120
DB 61 VSPPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKVAEYGVDCRLINPQINIGNEERIRM 122
QY 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDOFKVSV 180
DB 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDOFKVSV 182
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVSAIGLILSEGIGDTLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARGSAVSAIGLILSEGIGDTLRVSL 242
QY 241 AADPVEEIKVGFPIILKSIRSRGINPIACPTCSRQEPDVIGTVNALRQLEDITITPMV 300
DB 241 AADPVEEIKVGFPIILKSIRSRGINPIACPTCSRQEPDVIGTVNALRQLEDITITPMV 302
QY 301 SIIGCVNNGPGEALVSTLGVGNGKSGLYEDGVKRDLDNDMDIOLLEARIKASQOLD 359
DB 301 SIIGCVNNGPGEALVSTLGVGNGKSGLYEDGVKRDLDNDMDIOLLEARIKASQOLD 362
QY 360 DEARRIDVOQVEK 372
DB 360 DEARRIDVOQVEK 375
```

RESULT 7

H64063  
gcpb protein - Haemophilus influenzae (strain Rd KW20)

C/Species: Haemophilus influenzae  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #ext\_change 28-Jul-2000  
C/Accession: H64063  
R:Flotischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

Db 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALERLEDITITPMDV 300  
 Qy 301 SIIGCVNNGGEGALVSTLGTGKNGKSGLYEDGVRKRLDNNMDILEARIRAKASOLD 360  
 Db 301 SIIGCVNNGGEGALVSTLGTGKNGKSGLYEDGVRKRLDNNMDILEARIRAKASOLD 360  
 Qy 361 EARRIDVOQVEK 372  
 Db 361 EARRIDVOQVEK 372

## RESULT 2

A91051

hypothetical protein Ecac3377 [imported] - Escherichia coli (strain O157:H7, substrain R1  
 C/Species: Escherichia coli  
 C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001  
 C/Accession: A91051  
 R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene  
 A/Reference number: A9629; MUID:2156231; PMID:11258796  
 A/Accession: A91051  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-372 <NAV>  
 A/Cross-references: GB:BA000007; PIDN:BA936800.1; PID:g13362847; GSPDB:GN00154  
 A/Experimental source: strain O157:H7, substrain R1MD 0509952  
 C/Genetics:  
 A/Gene: Ecac3377  
 C/Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 2; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHNQAPIQRKSTRIVYGNVPIGDGAPFVQSMNTTRTTVEATVNOIKALERVGADIVR 60  
 Db 1 MHNQAPIQRKSTRIVYGNVPIGDGAPFVQSMNTTRTTVEATVNOIKALERVGADIVR 60  
 Qy 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCLRINPQNGNERIRM 120  
 Db 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCLRINPQNGNERIRM 120  
 Qy 121 VVDCARDKNIPIRIGVNASGLEKDLQEKYGEPTPOALLSARHVDHLRLNPDQKVS 180  
 Db 121 VVDCARDKNIPIRIGVNASGLEKDLQEKYGEPTPOALLSARHVDHLRLNPDQKVS 180  
 Qy 121 VVDCARDKNIPIRIGVNASGLEKDLQEKYGEPTPOALLSARHVDHLRLNPDQKVS 180  
 Db 121 VVDCARDKNIPIRIGVNASGLEKDLQEKYGEPTPOALLSARHVDHLRLNPDQKVS 180  
 Qy 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLLSBSIGDTRLVSL 240  
 Db 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLLSBSIGDTRLVSL 240  
 Qy 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALERLEDITITPMDV 300  
 Db 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALERLEDITITPMDV 300  
 Qy 301 SIIGCVNNGGEGALVSTLGTGKNGKSGLYEDGVRKRLDNNMDILEARIRAKASOLD 360  
 Db 301 SIIGCVNNGGEGALVSTLGTGKNGKSGLYEDGVRKRLDNNMDILEARIRAKASOLD 360  
 Qy 361 EARRIDVOQVEK 372  
 Db 361 EARRIDVOQVEK 372

## RESULT 3

E85895

hypothetical protein gcpe [imported] - Escherichia coli (strain O157:H7, substrain EDL93  
 C/Species: Escherichia coli  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C/Accession: E85895  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: AB5480; MUID:21074935; PMID:11206551  
 A/Accession: E85895  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-372 <STO>  
 A/Cross-references: GB:AE005174; NID:g12516909; PIDN:AA057625.1; GSPDB:GN00145; UWGP:237.  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: gcpe  
 C/Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 2; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHNQAPIQRKSTRIVYGNVPIGDGAPFVQSMNTTRTTVEATVNOIKALERVGADIVR 60  
 Db 1 MHNQAPIQRKSTRIVYGNVPIGDGAPFVQSMNTTRTTVEATVNOIKALERVGADIVR 60  
 Qy 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCLRINPQNGNERIRM 120  
 Db 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCLRINPQNGNERIRM 120  
 Qy 121 VVDCARDKNIPIRIGVNASGLEKDLQEKYGEPTPOALLSARHVDHLRLNPDQKVS 180  
 Db 121 VVDCARDKNIPIRIGVNASGLEKDLQEKYGEPTPOALLSARHVDHLRLNPDQKVS 180  
 Qy 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLLSBSIGDTRLVSL 240  
 Db 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSAGVSAIGLLSBSIGDTRLVSL 240  
 Qy 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALERLEDITITPMDV 300  
 Db 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALERLEDITITPMDV 300  
 Qy 301 SIIGCVNNGGEGALVSTLGTGKNGKSGLYEDGVRKRLDNNMDILEARIRAKASOLD 360  
 Db 301 SIIGCVNNGGEGALVSTLGTGKNGKSGLYEDGVRKRLDNNMDILEARIRAKASOLD 360  
 Qy 361 EARRIDVOQVEK 372  
 Db 361 EARRIDVOQVEK 372

## RESULT 4

AB0822

gcpe protein (protein E) [imported] - Salmonella enterica subsp. enterica serovar Typh  
 C/Species: Salmonella enterica subsp. enterica serovar Typh  
 A/Note: this species has also been called Salmonella typh  
 C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 18-Nov-2002  
 C/Accession: AB0822  
 R/Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connerton, P.; Croft, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moutle, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A/Reference number: AB0502; MUID:21534947; PMID:11677608  
 A/Accession: AB0822  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-372 <PAR>  
 A/Cross-references: GB:AL513382; PIDN:CAD02726.1; PID:g16503738; GSPDB:GN00176  
 C/Genetics:  
 A/Gene: STY2768  
 C/Superfamily: gcpe protein

Query Match 98.1%; Score 1831; DB 2; Length 372;  
 Best Local Similarity 98.1%; Pred. No. 3.5e-119;  
 Matches 365; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:07:18 ; Search time 26 Seconds  
(without alignments)  
1375.952 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866  
Sequence: 1 MHNQAPIQRKSTRIVYGVN.....RAKASQLEARRIDVQVEK 372

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	372	1 S23058	gcpe protein - Esc
2	1866	100.0	372	2 A91051	hypothetical prote
3	1866	100.0	372	2 E85895	hypothetical prote
4	1831	98.1	372	2 AB0822	gcpe protein (prot
5	1622.5	87.0	375	2 AG0350	probable acetyltra
6	1621.5	86.9	376	2 F82283	gcpe protein VC075
7	1603.5	85.9	368	1 H64063	gcpe protein - Hae
8	1372.5	73.6	371	2 F83171	conserved hypotet
9	1282	68.7	368	2 E84563	gcpe protein (impo
10	1042.5	55.9	383	2 H87354	gcpe protein (impo
11	887	47.5	349	2 G97121	gcpe protein (impo
12	882.5	47.3	367	2 AB3825	peptidoglycan acet
13	868	46.5	368	2 A11254	probable peptidogl
14	859.5	46.1	377	2 C69955	hypothetical prote
15	828	44.4	357	2 F70433	gcpe protein - Aqu
16	815.5	43.7	384	2 T35407	conserved hypotet
17	794	42.3	359	1 A64598	protein E - Helico
18	790	42.3	359	2 E71914	hypothetical prote
19	751.5	40.3	387	2 F70886	probable gcpe prot
20	712.5	38.2	392	2 G87106	conserved hypotet
21	706	37.8	344	2 D72321	gcpe protein - The
22	690	37.0	357	2 AB1339	gcpe protein homol
23	618.5	33.1	403	2 S77159	hypothetical prote
24	613.5	32.9	408	2 AF2118	probable gcpe prot
25	588	31.5	404	2 F71324	gcpe protein - Del
26	543.5	29.1	429	2 D75526	gcpe protein NMBl3
27	527	28.2	421	2 D81098	conserved hypotet
28	527	28.2	421	2 H81843	gcpe protein (impo
29	509	27.3	437	2 AH3285	

30	504.5	27.0	601	2 C81715	gcpe protein TC032
31	492	26.4	416	2 AB2911	peptidoglycan acet
32	492	26.4	440	2 H97685	hypothetical prote
33	487	26.1	602	2 E71562	hypothetical prote
34	470.5	25.2	613	2 C86537	gcpe protein (impo
35	470.5	25.2	613	2 E72087	gcpe protein CP038
36	459	24.6	417	2 B82542	conserved hypotet
37	126	6.8	533	2 S76101	hypothetical prote
38	114	6.1	524	2 H69099	conserved hypotet
39	112.5	6.0	686	2 H90226	minichromosome mai
40	111	5.9	486	2 A75045	2-isopropylmalate
41	108	5.8	690	2 E72337	translation initia
42	106.5	5.7	284	2 D86701	mathionine aminope
43	106.5	5.7	531	2 S52294	2-isopropylmalate
44	106.5	5.7	531	2 AH2410	2-isopropylmalate
45	106	5.7	748	2 A64319	carbon-monoxide de

## ALIGNMENTS

```
RESULT 1
S23058
gcpe protein - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #ext_change 01-Mar-2002
C/Accession: S23058; B65028
R/Baker, J.; Franklin, D.B.; Parker, J.
FEMS Microbiol. Lett. 94, 175-180, 1992
A/Title: Sequence and characterization of the gcpe gene of Escherichia coli.
A/Reference number: S23057
A/Accession: S23058
A/Molecule type: DNA
A/Residues: 1-372 <BAK>
A/Cross-references: EMBL:X64451, NID:g41540, PID:g41542
A/Experimental source: strain K-12
R/Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coj
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426517; PMID:9278503
A/Accession: B65028
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-372 <BAK>
A/Cross-references: GB:A6000338, GB:U00096, NID:g1788862, PIDN:AAC75568.1; PID:g1788863;
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: gcpe
C/Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 1; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIQRKSTRIVYGVNPIGDAPIAVGSMTRTDTVATNOIKALRVGADYR 60
DB 1 MHNQAPIQRKSTRIVYGVNPIGDAPIAVGSMTRTDTVATNOIKALRVGADYR 60
QY 61 VSPPTMDAAEAKLTKQGVNPLVADIHFDYRIALKVAEYGVDCRLNNGINERIRM 120
DB 61 VSPPTMDAAEAKLTKQGVNPLVADIHFDYRIALKVAEYGVDCRLNNGINERIRM 120
QY 121 VVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTPALLESAMRHVDHLDRLPQKVSU 180
DB 121 VVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTPALLESAMRHVDHLDRLPQKVSU 180
QY 122 VVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTPALLESAMRHVDHLDRLPQKVSU 180
DB 122 VVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTPALLESAMRHVDHLDRLPQKVSU 180
QY 181 KASDVFLAVESYRLAKOIDOPHLGITEAGAGSAGVSAIGLGLLSEGIQDLRVSL 240
DB 181 KASDVFLAVESYRLAKOIDOPHLGITEAGAGSAGVSAIGLGLLSEGIQDLRVSL 240
QY 241 AADPVEIKVGFIDILKSIRSRGINFIACPTCSQGEFVIGTVNALQORLEDITTPMDV 300
DB 241 AADPVEIKVGFIDILKSIRSRGINFIACPTCSQGEFVIGTVNALQORLEDITTPMDV 300
```

CC are used: (i) to increase the isoprenoid levels in viruses and cells;  
 CC (ii) for determining the enzymatic activity of gcpe and yf8b proteins;  
 CC and (iii) to identify compounds that inhibit activity of gcpe, i.e.  
 CC potential antibacterial, antimycotic, antiparasitic or antiviral agents  
 CC for use in humans or animals, or antiviral, antiparasitic, fungicidal  
 CC or herbicidal agents for agriculture.

XX Sequence 679 AA;

Query Match 24.7%; Score 460; DB 22; Length 679;  
 Best Local Similarity 38.9%; Pred. No. 2.3e-37;  
 Matches 112; Conservative 57; Mismatches 81; Indels 38; Gaps 8;

QY 9 RKKSTR-IYGNVPIGDGAPIAVQSMNTTRTVDVATVNOIKALEVAGDIYRVSVPTMD 67  
 DB 117 KRLPTREVVIGNVKIGNNKRIAIQTWASCDTRNVECVQIRCKDGLGADIVRLTVQGVQ 176  
 QY 68 AAEAF-----KLIKOQVNVPLVADIHFDYRIALKVAEYGVDCLRINPKNIGN----- 114  
 DB 177 EAQASHIKRKLSENNVNIPLVADIHFNPKIALMAADV-FEKIRVNGVYVDGRKKWIDK 235  
 QY 115 -----EERIRMVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALL 158  
 DB 236 VYKTEEPDEGKLFIEKEFVPLIEKCKR-LNRAIRIGTNHGLSSRVLSYGD-TPLATV 293  
 QY 159 ESAMRHVDHLDRNFPQFVSVKASDVFLAVESYRL-AKQIDQ----PLHIGITAGGA 213  
 DB 294 ESAMRFSDLCENNFNFLVFSMKASNAVYMIQSYRLVSKOYERNMMPFIHLGVTEAGFG 353  
 QY 214 RSGAVKSAIGLGLLSEIGDTRLVSLADPVEIKVGFILKSLRIR 261  
 DB 354 DNGRIKSYLGISGLYDGIQDITIRISLTEDPWEELTPCKKLVENLKKR 401

RESULT 15

AAW90883 ID AAW90883 standard; Protein; 752 AA.

XX AAW90883;

XX 07-JUL-2000 (first entry)

DE P. falciparum gcpe protein.

XX 1-deoxy-D-xylulose-5-phosphate reductoisomerase; isoprenoid biosynthesis;  
 KM deoxy-D-xylulose-phosphate; 1-deoxy-D-xylulose-5-phosphate synthase;  
 KW gcpe; antimycotic; antibiotic; antiviral; antiparasitic; antimicrobial;  
 KM bactericide; fungicide; herbicide; treatment; disease.

XX Plasmodium falciparum.

XX WO200017233-A2.

XX 30-MAR-2000.

XX 22-SEP-1999; 99WO-EP07055.

XX 22-SEP-1998; 98DE-1043279.

XX 21-MAY-1999; 99DE-1023567.

XX (JOMAA/) JOMAA H.

XX Jomaa H;

XX WPI; 2000-283543/24.

XX N-PDB; AAX82996.

PT New nucleic acid, useful therapeutically and to screen for e.g.  
 PT antimicrobials and herbicides, encode proteins involved in isoprenoid  
 PT biosynthesis by the deoxy-D-xylulose-phosphate route

PS Claim 3; Page 41-44; 44pp; German.

CC This invention describes novel polypeptides and their analogs which are  
 CC involved in isoprenoid biosynthesis by the deoxy-D-xylulose-phosphate  
 CC route. The proteins described include 1-deoxy-D-xylulose-5-phosphate  
 CC reductoisomerase, 1-deoxy-D-xylulose-5-phosphate synthase and the gcpe  
 CC protein. The products of the invention have antimycotic, antibiotic,  
 CC antiviral, antiparasitic, antimicrobial, bactericide, fungicide and  
 CC herbicide activity. The encoding nucleic acid sequences of the invention  
 CC are used for prevention or treatment of diseases in humans and animals  
 CC and also to screen compounds enzyme inhibitory activity (potential  
 CC antimicrobials, antiparasitic agents, antivirals, fungicides,  
 CC bactericides and herbicides, for use in human or veterinary medicine or  
 CC agriculture). This sequence represents the Plasmodium falciparum gcpe  
 CC protein described in the method of the invention.

XX Sequence 752 AA;

Query Match 23.9%; Score 446; DB 21; Length 752;  
 Best Local Similarity 38.2%; Pred. No. 7.1e-36;  
 Matches 110; Conservative 56; Mismatches 84; Indels 38; Gaps 8;

QY 9 RKKSTR-IYGNVPIGDGAPIAVQSMNTTRTVDVATVNOIKALEVAGDIYRVSVPTMD 67  
 DB 117 KRLPTREVVIGNVKIGNNKRIAIQTWASCDTRNVECVQIRCKDGLGADIVRLTVQGVQ 176  
 QY 68 AAEAF-----KLIKOQVNVPLVADIHFDYRIALKVAEYGVDCLRINPKNIGN----- 114  
 DB 177 EAQASHIKRKLSENNVNIPLVADIHFNPKIALMAADV-FEKIRVNGVYVDGRKKWIDK 235  
 QY 115 -----EERIRMVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALL 158  
 DB 236 VYKTEEPDEGKLFIEKEFVPLIEKCKR-LNRAIRIGTNHGLSSRVLSYGD-TPLGMY 293  
 QY 159 ESAMRHVDHLDRNFPQFVSVKASDVFLAVESYRL-AKQIDQ----PLHIGITAGGA 213  
 DB 294 ESAMRFSDLCENNFNFLVFSMKASNAVYMIQSYRLVSKOYERNMMPFIHLGVTEAGFG 353  
 QY 214 RSGAVKSAIGLGLLSEIGDTRLVSLADPVEIKVGFILKSLRIR 261  
 DB 354 DNGRIKSYLGISGLYDGIQDITIRISLTEDPWEELTPCKKLVENLKKR 401

Search completed: November 23, 2003, 17:15:04  
 Job time : 53 secs

Db 258 GIIKSAVIGITLAEGLDITRCSLTGCPTEIRVDCSLNHTX 302

## RESULT 13

AAE19650 standard; Protein; 740 AA.

AAE19650;

31-MAY-2002 (first entry)

Arabidopsis thaliana GCPE protein.

GCPE gene: methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate; transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate; food; feed source; transfection; single nucleotide polymorphism; SNP; oxidative stress tolerance; UV tolerance; transformation; chromosome V; plant; GCPE protein.

Arabidopsis thaliana.

Key Location/Qualifiers

Cleavage-site 32..33 /note= "Plastidial targeting peptide cleavage site"

MO200212478-A2.

14-FEB-2002.

06-AUG-2001; 2001MO-US24335.

07-AUG-2000; 2000US-223483P.

(MONS ) MONSANTO TECHNOLOGY LLC.

Boronat A, Campos N, Rodriguez-concepcion M, Rohner M, Seeman M;

Valentin HE, Venkatesh TV, Venkatramesh M;

WPI: 2002-227151/28.

N-PSDB; AAD31201.

GCPE nucleic acid which is an essential gene of the methyl-D-erythritol phosphate pathway, encoding a fully defined GCPE protein which is useful for increasing levels of tocopherol substrates in plants

Claim 1; Page 140-142; 155pp; English.

The invention relates to gcpe nucleic acid molecule, an essential gene of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice, Arabidopsis thaliana or Escherichia coli GCPE protein. gcpe is useful for producing a transgenic plant such as Brassica campestris, B. napus, canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard, oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower, or wheat with an increased isoprenoid (tocopherol) compound level. The expression of GCPE protein in organisms increases the level of tocopherol substrate such as isopentyl diphosphate and dimethylallyl diphosphate biosynthesis. Transgenic organisms overexpressing GCPE protein can nutritionally enhance food and feed sources. Overexpression of GCPE protein in transgenic plant may provide tolerance to stresses e.g., oxidative stress tolerance such as to oxygen or ozone, UV tolerance, etc. GCPE may be used to obtain nucleic acid molecules from the same species, and to obtain nucleic acid homologues. gcpe is also used as or primers. The recombinant vectors are used in plant transformation or transfection. gcpe an also act as markers capable of detecting polymorphisms such as single nucleotide polymorphisms (SNPs). GCPE is also used to determine the level or pattern of expression of the protein. The present sequence is Arabidopsis thaliana GCPE protein. gcpe gene is located on chromosome V.

Sequence 740 AA;

Query Match 24.8%; Score 462; DB 23; Length 740;

Best Local Similarity 38.6%; Pred. No. 1.6e-37; Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RKSTRIYGNVPIGDPAPAVOSMTNRTTVDVATVNOIKALERVADIVRSVPTMDA 68  
Db 84 RKRTRVWGNVNLGSEHPRIQITMTSDPKDITGVDEVWRIADGADIVRTVQKKE 143  
QY 69 AEAFA-----KLIRQGVNVPVADIDHFDYRIALVAEYGVDCLRINQINIGERIRRVVD 123  
Db 144 ADACFEIKDVLVOLANNIPVADIHFAPTVALVAEC-PDKIVNNGNFPADRAQFETID 202  
QY 124 CARDK-----NIPRIGVAGSLKXDLQEKYGEPTQALLLSA 161  
Db 203 YTEDEVQKELQHIQVFTPLVEKCKKYGAMRIGTNHGSLSDRIMSYGD-SPRGWESA 261  
QY 162 MRVVDHLDRINFPQFVKSVASDVFLAVESYRLAAQI-----DPLHLGITAGARSG 216  
Db 262 FEPARIQRKLDYNFVFSMKASNPVIMVOYRLLVEMVHGWDYPLHLGVTBAGEGEGD 321  
QY 217 AVKSAIGLGLLSGIGDTRVSLADPVEEI 248  
Db 322 RKSATIGITLQDGLDITRVSITPPEEI 353

## RESULT 14

AAE45693 standard; Protein; 679 AA.

AAE45693;

15-MAR-2001 (first entry)

P. falciparum gcpe protein.

Isoprenoid; gcpe; yfgb; antimicrobial; transgenic plant; agriculture; antimycotic; antiparasitic; antiviral; fungicidal; herbicidal.

Plasmodium falciparum.

MO200072022-A1.

30-NOV-2000.

20-MAY-2000; 2000MO-EP04592.

21-MAY-1999; 99DE-1023567.

21-MAY-1999; 99DE-1023568.

(JOMA/) JOMAA H.

Jomaa H;

WPI: 2001-025196/03.

N-PSDB; AAC82654.

Incorporating gcpe and yfgb genes into viruses and cells, for increasing isoprenoid content and identifying e.g. antimicrobial agents, comprises using DNA sequences from bacteria or parasites

Disclosure; Page 21-23; 36pp; German.

This invention describes a novel method for incorporating gcpe and yfgb genes into viruses and cells for increasing isoprenoid content and identifying e.g. antimicrobial agents, comprises using DNA sequences (I) from the gcpe or yfgb genes of bacteria or parasites or DNA sequences (II) which hybridize to the specified genes or encode a plastid protein with the same biological activity as those encoded by the genes. The invention also describes (1) plant cells containing (I) or (II); (2) transformed plant cells, and transgenic plants regenerated from them, that contain (I) or (II); (3) determining the enzymatic activity of a gcpe protein; or (4) screening compounds (A) that have antimycotic, antiparasitic or antiviral activity in humans or animals or antiviral, antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)





CC encoded by AB271062 to AB271130 (II), which are isolated from  
CC Mycobacterium tuberculosis. (I) are encoded by genes (II) whose  
CC expression is induced or up-regulated during culture of a mycobacterium  
CC under conditions defined by a dissolved oxygen tension of at least 10%  
CC air saturation measured at 37 plus degrees Celsius, when compared with a  
CC dissolved oxygen tension of at least 40% air saturation measured at 37  
CC plus degrees Celsius. (I) and (II) have antibacterial, tuberculostatic  
CC and immunostimulant activities, and can be used in vaccines and gene  
CC therapy. (I) and (II) can be used for the manufacture of a medicament  
CC for treating or preventing a mycobacterial infection. They can also be  
CC used for the manufacture of a diagnostic reagent for identifying a  
CC mycobacterial infection.

XX Sequence 387 AA;

Query Match 40.3%; Score 751.5; DB 24; Length 387;  
Best Local Similarity 42.7%; Pred. No. 2e-67;  
Matches 156; Conservative 64; Mismatches 134; Indels 11; Gaps 2;

QY 9 RRRKSTRIVGVNPIGDGAPIAVOSMTNTRTTDVEATVNOIKALERVGADIVRSVPTMDA 68  
DB 18 RRAATROLVGNVGVSGDHPVSVOSMCTTKTHDVNSTIQIALAELTAGCDIVRVACPRQED 77  
QY 69 AEAFLIKQOVNPLVADHFDVRIALKVAEYGVDCIRINPNIQN-EEBIRMVDCARD 127  
DB 78 ADALAEIARHSQIPVVDIHFQPRYIFPAIDAGCAAVRNPNIKEFDGRVGEVAKAAGA 137  
QY 128 KNPIPIRGVNVGSLKLOEKKEPTPOALLSESMRVHDLNLPQFVSVKASVFL 187  
DB 138 AGPIPIRGVNVGSLDKRMEKYGATPEALVESALLESFEGFDIKISVGHNDPVV 197  
QY 188 AVESYRLAKQIDQPLHIGTEAGARSAGVAKSAIGLLSEGIGDTLRVSLAADPVEE 247  
DB 198 MVAAYELLARCVPLHIGTEAGPARQGTIKSAVAFALLSGIGTITVLSAPVVEE 257  
QY 248 IKTGFILKSLRIRSRGINFIACTPCTSRQEFVIGTVNALEQRLIEDITPMDSIIGCVV 307  
DB 258 VKVGNVLESLNRPRLSEIVSCPSCGRAQVDVTLANEVTLAGLDGDLVRLVAVMGCVV 317  
QY 308 NGEGEALVSTLTGTGNGKSGLYEDGVKRLDNNDMIDLEARIKAKSQLEARRIDV 367  
DB 318 NGGEAREADLVASGNGKQITVRGEVITKTPBEAQIVETL-----IEAMRLAA 367  
QY 368 QQVEK 372  
DB 368 EMGEQ 372

RESULT 10

ABP78786  
ID ABP78786 standard; Protein; 421 AA.

XX ABP78786;

XX 07-MAR-2003 (first entry)

DE N. gonorrhoeae amino acid sequence SEQ ID 4102.

XX Antibacterial; infection; vaccine; gene therapy.

OS Neisseria gonorrhoeae.

XX WO200279243-A2.

XX 10-OCT-2002.

XX 12-FEB-2002; 2002WO-IB02069.

XX 12-FEB-2001; 2001GB-0003424.

XX (CHIR-) CHIRON SPA.

XX Fontana MR, Pizsa M, Masignani V, Monaci E;

XX WPI; 2003-058415/05.  
DR N-PDB; AB239756.

PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
XX medicament for treating or preventing N. gonorrhoeae infection

PS Disclosure; Page 487; 815pp; English.

CC The present invention relates to proteins from Neisseria gonorrhoeae.  
CC Also disclosed are the nucleic acid molecules encoding the proteins and  
CC antibodies that specifically bind to the proteins. The composition  
CC comprising the protein, nucleic acid or antibody is useful for the  
CC manufacture of a medicament for treating or preventing N. gonorrhoeae  
CC infection, this may be in the form of a vaccine or gene therapy.  
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX molecules of the invention.

XX Sequence 421 AA;

Query Match 28.6%; Score 534; DB 24; Length 421;  
Best Local Similarity 32.3%; Pred. No. 3.1e-45;  
Matches 135; Conservative 87; Mismatches 142; Indels 54; Gaps 12;

QY 7 IORRSTRIVGVNPIGDGAPIAVOSMTNTRTTDVEATVNOIKALERVGADIVRSVPTMDA 66  
DB 4 LORRKHQVLDHITVGSAPVVGSMNTTDTADAKATLMOIKELSDASSEWRTIVNSP 63  
QY 67 DAAEAFLIKQOVN-----VPLVADHFD-VRIALKVAEYGVDCIRINPNIQN-EEBIRMVDCARD 114  
DB 64 EAAASKVAEIRRRLLDVGAVTPLIGDFHNGERLAEFFPCGKALSKYRINPVGKGVKG 123  
QY 115 EEBIRMVDCARDKNIPRIPIRGVNVGSLKLOEKKEPTPOALLSESMRVHDLNLPQFVSVKASVFL 161  
DB 124 DEKPAFMITTALENDGAVRIGVNWSLDSQSLAKRMMDANLVASAPRPEEVNKEALIVSA 183  
QY 162 MRHVHDLNLPDOFV--SVKASDVFLAVESYRLAKQIDQPLHIGTEAGARSAGVAK 219  
DB 184 LESAERAVVLLGPEDKILISCKVSAVHDLIOYRRELGSVCVPLHIGLEAAGMSGIYA 243  
QY 220 SAIGLLSEGIGDTLRVSLAADP-----VEEIKVGFILKSLRIRSRGINFIACTPCTSR 275  
DB 244 STAAISVLLQEGIGDTIRISLTPPCSPRTQVGVVGOEITLGMGLRSPFPMVTACPGCGR 303  
QY 276 QEFVIGTV-----NALER-----LEDTITPMDSIIGCVVNGEGEALVSTLGV---- 320  
DB 304 TTSVTFQELADQVQVYLRQKMSIMRTLYPGVESLNVAVMGCVVNGEGESKLADIGISLPG 363  
QY 321 TCGNKKSGLYEDGVKRLDNNDMIDLEARI-----AKASQLEARRIDVQOV 370  
DB 364 TGETPVAPVYVDGERKVTLKGNINASEFLAIVEEYKTVNGKSSKRKNGKVIPIOSL 421

RESULT 11

AA37164  
ID AA37164 standard; Protein; 611 AA.

XX AA37164;

XX 07-OCT-1999 (first entry)

DE Amino acid sequence of a Chlamydia trachomatis protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;  
XX paratrachoma; inclusion conjunctivitis; genital disease; perhepatitis;  
XX nongonococcal urethritis; cervicitis; salpingitis;  
XX Bartholinitis; pneumonia; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

QY 69 AAEFKIKQGVNPLVADHFDYRIALKAAYGVDCLRINPNIEN-EEIRIMVVDCAAD 127  
 Db 65 AEAFLPIIAKSSPIPIADIHFQPKYIFPAIDAGCAAVRPNIGIKEDPKGVKAAAGD 124  
 QY 128 KNIPRIGVAGSLEKDLOEKY-GEPIPOLLSANRHHVDRLNFDQKVSFKASDVF 186  
 Db 125 AGPIPIGVAGSLEKDLOEKY-GEPIPOLLSANRHHVDRLNFDQKVSFKASDVF 184  
 QY 187 LAEVSRYLLAKQIDOPHLGITEAGARGAVKSAIGLLSEGIGDTLRVSLAADPVE 246  
 Db 185 LMEVETRLQLEQSDYPLHLGVTEAGRPKFTIKSSVAFGLLSOGIGDTLRVSLAADPVE 244  
 QY 247 EIKVGFILKSLIRKSGINFIACPTCSROEPVIGTVNALBQRLDITIPMDVSIIGCV 306  
 Db 245 EIKVGFILKSLIRKSGINFIACPTCSROEPVIGTVNALBQRLDITIPMDVSIIGCV 304  
 QY 307 VNGPGEALVSTLGVTKGNKSGLYEDGVKRDRLDNDMDIQLEARIKAKASQLEARRI 365  
 Db 305 VNGPGEALVSTLGVTKGNKSGLYEDGVKRDRLDNDMDIQLEARIKAKASQLEARRI 363

## RESULT 8

AA81205  
 ID AA81205 standard; Protein; 387 AA.

AA81205;

04-SEP-2001 (first entry)

Mycobacterium tuberculosis potential drug target protein SEQ ID 256.

Drug target; growth; organism viability; characterisation.

Mycobacterium tuberculosis.

MO200135317-A1.

17-MAY-2001.

13-NOV-2000; 2000MO-US31152.

12-NOV-1999; 99US-0165086.

12-NOV-1999; 99US-0165124.

01-FEB-2000; 2000US-0179531.

(REGC ) UNIV CALIFORNIA.

Eisenberg D, Rotstein SH, Marcotte EM,

WPI; 2001-329193/34.

N-PSDB; AAH52056.

Identifying nucleotide or polypeptide sequence for use as drug target,

involves providing algorithm that analyzes a functional relationship

between nucleotide or polypeptide sequences, and comparing the

sequences

Disclosure; Page 182; 207pp; English.

This invention relates to a method for identifying a nucleotide or

polypeptide sequence that may be a drug target, or essential for growth

or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092

represent DNA encoding proteins AA81096 - AA81241. Mycobacterium

tuberculosis proteins which are potential drug targets. The DNA and

protein sequences are used to illustrate the method of the invention. The

method involves providing an unknown nucleotide or polypeptide sequences,

and comparing it to a number of sequences along with at least one

algorithm capable of analyzing a functional relationship between

nucleotide and polypeptide sequences. The method is useful for

characterising the function of nucleic acids and polypeptides that may be

useful as a target for a drug or essential for the growth or viability of

an organism.

Sequence 387 AA;

Query Match 40.3%; Score 751.5; DB 22; Length 387;

Best Local Similarity 42.7%; Pred No. 26-67;

Matches 166; Conservative 64; Mismatches 134; Indels 11; Gaps 2;

QY 9 RKRSTRIYVGNVPIGDGAPIAVQSMNTTTRTVEATVQIKALERVADIVRSVPTMDA 68  
 Db 18 RATRQLVMGVNGVGDHVSQVSMCTKTHTDYNSTLQIAELTAAACDIIVRVACPRQED 77  
 QY 69 AAEFKIKQGVNPLVADHFDYRIALKAAYGVDCLRINPNIEN-EEIRIMVVDCAAD 127  
 Db 78 AALAEIASHSQIPVADHFDYRIALKAAYGVDCLRINPNIEN-EEIRIMVVDCAAD 137  
 QY 128 KNIPRIGVAGSLEKDLOEKY-GEPIPOLLSANRHHVDRLNFDQKVSFKASDVF 187  
 Db 138 AGPIPIGVAGSLEKDLOEKY-GEPIPOLLSANRHHVDRLNFDQKVSFKASDVF 197  
 QY 188 AAEFKIKQGVNPLVADHFDYRIALKAAYGVDCLRINPNIEN-EEIRIMVVDCAAD 247  
 Db 198 MVAAYEELIARCDYPLHLGVTEAGRPKFTIKSSVAFGLLSOGIGDTLRVSLAADPVE 257  
 QY 248 EIKVGFILKSLIRKSGINFIACPTCSROEPVIGTVNALBQRLDITIPMDVSIIGCV 307  
 Db 258 VIKVGNVLESILNIRPSLEIVSCPCSGRAQVYVTLANVTAGLDLDPRLVAVMGCV 317  
 QY 308 NGPGEALVSTLGVTKGNKSGLYEDGVKRDRLDNDMDIQLEARIKAKASQLEARRI 367  
 Db 318 NGPGEALVSTLGVTKGNKSGLYEDGVKRDRLDNDMDIQLEARIKAKASQLEARRI 367  
 QY 368 QOVER 372  
 Db 368 EMGEQ 372

## RESULT 9

ABP57500  
 ID ABP57500 standard; Protein; 387 AA.

ABP57500;

28-APR-2003 (first entry)

Mycobacterium tuberculosis protein SEQ ID NO:129.

Mycobacterium tuberculosis; infection; antibacterial; tuberculostatic;

KW immunostimulant; vaccine; gene therapy; mycobacterial infection.

Mycobacterium tuberculosis.

WO2003000721-A2.

03-JAN-2003.

21-JUN-2002; 2002MO-GB02845.

22-JUN-2001; 2001GB-0015365.

07-SEP-2001; 2001GB-0021780.

(MICR-) MICROBIOLOGICAL RES AUTHORITY.

James BW, Bacon J, Marsh P;

WPI; 2003-201403/19.

N-PSDB; ABZ71126.

New mycobacterial peptide, its fragment, variant or derivative, useful

as vaccine for treating or preventing mycobacterial infections, and as

diagnostic reagents for identifying such infections

Claim 2; Page 235-236; 246pp; English.

ABP57436 to ABP57504 represent mycobacterial amino acid sequences (1)

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KM dermatological; osteopathic; neuroprotectant.  
XX  
OS Propionibacterium acne.  
XX  
PN MO200181581-A2.  
XX  
PD 01-NOV-2001.  
XX  
PF 20-APR-2001; 2001MO-US12865.  
XX  
PR 21-APR-2000; 2000US-199047P.  
XX  
PR 02-JUN-2000; 2000US-208841P.  
XX  
PR 07-JUL-2000; 2000US-216747P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Skeelzy YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'atsomeuve J, Zhang Y, Jen S, Carter D;  
XX  
XX WPI: 2001-616774/71.  
DR N-PSDB; AASS9573.  
XX  
XX Propionibacterium acne polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris -  
XX  
XX Example 1; SEQ ID NO 17407; 10699P; English.  
XX  
XX Sequences AAU39105-AAU68017 represent propionibacterium acne immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acne. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acne is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acne in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acne proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acne polypeptides and  
CC therefore treat P. acne infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acne presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 415 AA:  
SQ  
Query Match 42.1%; Score 785.5; DB 22; Length 415;  
Best Local Similarity 44.2%; Pred. No. 7.9e-71;  
Matches 159; Conservative 68; Mismatches 132; Indels 1; Gaps 1;  
QY 9 RRRSTIYGVNPIGCGAPPAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTMDA 68  
DB 53 RRRTHOIKVGDVIVGSGAPISVOSMTTKTHNDGATLQOIAALTAAGCDIVRAVACPTDXD 112  
QY 69 AEAFTKIKQOVNPLVADHFDVRIALKAVEGVDCRINPGIGN-EERIIRVNVCCARD 127  
DB 113 AEVLPIIAKSGQIPVADHFDVRIALKAVEGVDCRINPGIGN-EERIIRVNVCCARD 172  
QY 128 KNPIRIGVAGSLKEDLOEKYPTPOALLSASMBHVHDLRNFDOFVSVKASDVPL 187  
DB 173 HGHSIIRIGVAGSLKEDLOEKYPTPOALLSASMBHVHDLRNFDOFVSVKASDVPL 232  
QY 188 AVESYLLAKQIDOPHLGITEGARSAGVSAIGLLSGIGDITLRSVLAADPVE 247  
DB 233 MRAVEQLAAKCDYPLHLGTEGAPFOGRTKSAVAFGHLLAGIDITRVSISADPVE 292

QY 248 IKVGFPIKSLRISRGINFACPTCSROEPFDVIGTVNALGEORLEDIITPMDSIIGCVV 307  
DB 293 VKVGIKILSLNLRPRGLIEVSCPCRCQVDVLTLANVDYALBEGIDAPLRYAVMGCVV 352  
QY 308 NGPGEALVSTLGTGKNSGLYEDGVKRDLDNNDIMIQLEARIKAKSOLDDEARIDV 367  
DB 353 NGIGEGREADLVGAAGNGKGIKFKHGEVIRTVPEGEIVOLFVGEANRMADEMDDTGADEV 412  
RESULT 7  
AAG91951  
ID AAG91951 standard; Protein, 378 AA.  
XX  
XX AAG91951;  
AC  
XX  
DT 26-SEP-2001 (first entry)  
XX  
XX C glutamicum protein fragment SEQ ID NO: 5705.  
XX  
XX Corynebacterium; amino acid synthesis; vitamin; saccharide;  
KM organic acid synthesis.  
XX  
XX Corynebacterium glutamicum.  
XX  
XX BP1108790-A2.  
PN  
XX  
PD 20-JUN-2001.  
XX  
XX 18-DEC-2000; 2000EP-0127688.  
PF  
XX  
XX 16-DEC-1999; 99GP-0377484.  
PR 07-APR-2000; 2000JP-0159162.  
XX  
XX 03-AUG-2000; 2000JP-0280988.  
XX  
XX (KYOW ) KYOWA HAKKO KOGYO KK.  
PA  
XX  
XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;  
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;  
XX  
XX WPI: 2001-376931/40.  
DR N-PSDB; AAH67170.  
XX  
XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene  
PT  
XX  
XX Claim 17; SEQ ID NO: 5705; 246pp + Sequence Listing; English.  
PS  
XX  
XX The present invention provides a number of nucleotide and protein  
CC sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of corynebacterium bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Corynebacterium bacterium, and identifying a homologue of a gene derived  
CC from corynebacterium bacterium. Corynebacterium bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a protein described  
CC in the exemplification of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.  
XX  
XX Sequence 378 AA:  
SQ  
Query Match 40.8%; Score 761; DB 22; Length 378;  
Best Local Similarity 44.3%; Pred. No. 2.1e-68;  
Matches 159; Conservative 64; Mismatches 124; Indels 12; Gaps 3;  
QY 9 RRRSTIYGVNPIGCGAPPAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTMDA 68  
DB 5 RRRTHOIKVGDVIVGSGAPISVOSMTTKTHNDGATLQOIAALTAAGCDIVRAVACPTDXD 64

CC given in ABP65258 to ABP6354 ligated in frame to a polynucleotide  
 CC encoding a heterologous polypeptide. (1) has anti-diarrheic and  
 CC antibacterial activities, and can be used as an inhibitor of *Salmonella*.  
 CC (1) (which is a probe) is useful for the detection and/or identification  
 CC of *Bifidobacterium longum* in a biological sample. A carrier containing  
 CC the lactic acid bacterium *Bifidobacterium longum* NCC2705 (CNCM I-2618)  
 CC can be used for preventing and/or treating diarrhoea brought about by  
 CC pathogenic bacteria and/or rotavirus. The carrier is a food composition  
 CC selected from milk, yogurt, curd, cheese, fermented milks, milk based  
 CC fermented products, ice-creams, fermented cereal based products, milk  
 CC based powders, infant formula, pet food or a pharmaceutical composition  
 CC selected from tablets, liquid bacterial suspensions, dried oral  
 CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.  
 CC (1) is useful in DNA arrays or chips to carry out analysis of the  
 CC expression of the *Bifidobacterium* gene. AB081844 to AB081850 represent  
 CC *Bifidobacterium* related nucleotide sequences given in the Sequence  
 CC listing from the present invention but not mentioned further within the  
 CC specification.  
 CC N.B. The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied by the  
 CC European Patent Office.

XX Sequence 403 AA;

Query Match 43.6%; Score 814.5; DB 23; Length 403;  
 Best Local Similarity 46.3%; Pred. No. 8.4e-74; Indels 13; Gaps 4;  
 Matches 167; Conservative 70; Mismatches 111;

QY 3 NOAPIQ-RRKSTRIVYGNVPIGDAPIAVOSMTNTRTTDVEATVNOIKALERVGADIVRV 61  
 DB 22 SESPLHRRKSRRLIMVGPVGGAPISVQSMNTTLLANVPATLQIAELTAAGCDIVRV 81  
 QY 62 SVPTMDAAEFKLIKQVNVPLVADHFDYRIALKYAEVGDCLRNPGNIGNEERRM\_120  
 DB 82 AVBSODDADALPEICRKSPIPIADHIFOSKYVQALDACAARVNPGRIRKFEDEGPD 141  
 QY 121 VVOCARDKINPIRIGVAGSLKQLOEKYGEPTPOALLSARHVDHLDLNFDPQVSV 180  
 DB 142 ICAATDAGISLRIGVAGSLKQLOEKYGEPTPOALLSARHVDHLDLNFDPQVSV 201  
 QY 181 KASDVLAVESYRLAKQIDOPHLGITTEAGSAGVKSAGLGLLSEIGDITLAVSL 240  
 DB 202 KKHVDIVMTETRYRLAKSGWPHLGTTEAGSAGVKSAGLGLLSEIGDITLAVSL 261  
 QY 241 AADPVEIKTGFILKSLRIRSGINFIACPTCSROFEDVIGTVNLEQRLIEDITPMDV 300  
 DB 262 SAPPABEVKVGCKLEWTGMRPRKFDIISCPSCGACQAVDVIQLASAVTEGLKQVTPAIRV 321  
 QY 301 SIIGCVNNGPGBALVSTLGVTKGKKSGLYEDG-----VRKDRLDN-----NDMIDOLE 349  
 DB 322 AVMGCIYNGPGBEARADLVGASNGKQIFTKGVITVPEBDQIVDTLLITANDIAAOME 381  
 QY 350 A 350  
 DB 382 A 382

RESULT 5  
 AAM98298  
 ID AAM98298 standard; Protein; 359 AA.  
 XX  
 AC AAM98298;  
 XX  
 DT 31-MAR-1999 (first entry)  
 XX  
 DE H. pylori GHPO 76 protein.  
 XX  
 KM GHPO protein; *Helicobacter* infection; gastroduodenal disease; gastritis;  
 XX  
 OS peptic ulcer disease.  
 XX  
 XX *Helicobacter pylori*.  
 XX  
 XX MO9843478-A1.

XX  
 PD 08-OCT-1998.  
 XX  
 PF 01-APR-1998; 98WO-US06371.  
 XX  
 XX 29-JUL-1997; 97US-0902615.  
 PR 01-APR-1997; 97US-0833457.  
 PR 24-JUN-1997; 97US-0881227.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.  
 DR N-PSDB; AAX14017.  
 XX  
 XX Al-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;  
 DR WPI; 1998-542293/46.  
 XX  
 XX N-PSDB; AAX14017.  
 XX  
 PT New isolated *Helicobacter* polynucleotides - used to develop products  
 PT for the diagnosis, prevention and treatment of *Helicobacter*  
 PT infections and gastrointestinal diseases  
 XX  
 PS Claim 8; Page 372-374; 2054pp; English.

CC This sequence represents a *Helicobacter pylori* GHPO protein of the  
 CC invention. The polypeptides can be used for preventing or treating  
 CC *Helicobacter* infections, and gastroduodenal diseases associated with  
 CC these infections, including acute, chronic, and atrophic gastritis, and  
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be  
 CC used for the production of antibodies. The products can also be used for  
 CC detection and diagnosis.

SO Sequence 359 AA;

Query Match 42.6%; Score 794; DB 19; Length 359;  
 Best Local Similarity 45.8%; Pred. No. 8.5e-72; Indels 2; Gaps 2;  
 Matches 162; Conservative 73; Mismatches 117;

QY 9 RRSKSTRIVYGNVPIGDAPIAVOSMTNTRTTDVEATVNOIKALERVGADIVRVPTMDA 68  
 DB 5 RVKTKQIFIGVAGIIGDAPISVQSMNTTLLANVPATLQIAELTAAGCDIVRV 64  
 QY 69 AEAFLIKQVNVPLVADHFDYRIALKYAEVGDCLRNPGNIGNEERIMVVDCAEDK 128  
 DB 65 ALAKELKQVNVPLVADHFDYRIALKYAEVGDCLRNPGNIGNEERIMVVDCAEDK 123  
 QY 129 NIPRIGVAGSLKQLOEKYGEPTPOALLSARHVDHLDLNFDPQVSVKASVFLA 188  
 DB 124 NIPRIGVAGSLKQLOEKYGEPTPOALLSARHVDHLDLNFDPQVSVKASVFLA 182  
 QY 189 VESYRLAKQIDOPHLGITTEAGSAGVKSAGLGLLSEIGDITLAVSLAADPVEEI 248  
 DB 183 IEAYRMLRPLVITVPFLGTTEAGSAGVKSAGLGLLSEIGDITLAVSLAADPVEEI 242  
 QY 249 KVGFDILKSLRIRSGINFIACPTCSROFEDVIGTVNLEQRLIEDITPMDVSIICVNV 308  
 DB 243 KVABALIRSGIRKESINWISCTCGRIANLVDMIAIKVEXKSHIKTLPDISVMCCVNV 302  
 QY 309 GPGBALVSTLGVTKGKKSGLYEDGVKRDLDNNDMIDOLEARIRAKASQOLDEA 362  
 DB 303 ALGEAKHADMALAFGNRSGIIEKGVIIHKLAEKDFETFEVLEVMIAKERERS 356

RESULT 6  
 AAU56212  
 ID AAU56212 standard; Protein; 415 AA.  
 XX  
 AC AAU56212;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #17108.  
 XX  
 XX SAPHO syndrome; synovitis; acne; pustulosis; hypertostis; osteomyelitis;  
 XX

QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD 360  
DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD 360  
QY 361 EARRIDVQVQVEK 372  
DB 361 EARRIDVQVQVEK 372

RESULT 3  
ABBA7958  
ID ABBA7958 standard; Protein, 368 AA.  
AC ABBA7958;  
XX  
XX 05-FEB-2002 (first entry)  
DE Listeria monocytogenes protein #662.  
XX  
XX Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;  
KM vitamin B12; bacterial infection; disease.  
XX  
OS Listeria monocytogenes.  
XX  
XX WO200177335-A2.  
XX  
XX 18-OCT-2001.  
XX  
XX 11-APR-2001; 2001WO-FR01118.  
XX  
XX 11-APR-2000; 2000FR-0004629.  
XX  
XX (INSP ) INST PASTEUR.  
XX  
PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihl H, Dehoux P,  
PI Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P,  
PI Daniels J, Goebel W, Krefz J, Kuhn M, Ng E, Vazquez-Boland JA,  
PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,  
PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L,  
PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,  
PI Madueno E, De Pablos B, Weiland J, Kaerst U, Entian K, Hauf J,  
PI Rose M, Voss H;  
XX  
XX MPI; 2002-010914/01.  
XX  
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment  
PT and prevention of Listeria and related bacterial infections, and  
PT related polypeptides -  
XX  
XX  
XX Claim 6; SEQ ID No 663; 192pp; French.

CC The present invention relates to the genome sequence of Listeria  
CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of  
CC it are useful for selecting probes and primers for detecting genes in L.  
CC monocytogenes and related organisms, and for studying genetic  
CC polymorphisms and other genomes. The present sequence is a protein  
CC encoded by the genome sequence of the present invention. Proteins  
CC expressed from the genome sequence are useful for raising specific  
CC antibodies, identification of L. monocytogenes and related organisms, and  
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin  
CC B12. The genome sequence and proteins encoded by it are also useful for  
CC selecting compounds that regulate gene expression and cell replication  
CC and modulate L. monocytogenes-related diseases. In addition, the genome  
CC sequence and proteins encoded by it are useful in pharmaceutical and  
CC vaccines compositions for the treatment or prevention of infections by L.  
CC monocytogenes and related organisms.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIP0  
CC at ftp.wip0.int/pub/published\_pct\_sequences.

XX  
XX Sequence 368 AA;  
XX  
XX Query Match 46.5%; Score 868; DB 23; Length 368;

Best Local Similarity 49.2%; Pred. No. 2.6e-79;  
Matches 178; Conservative 67; Mismatches 115; Indels 2; Gaps 2;  
QY 10 RKSTR-IYVGNPFIGGAPLAVQSMRTTDTVEATVNOIKALERGADIVRSVPTMDA 68  
DB 7 RENTRPVQVGNLTIGSEBELTIGSMTTTTHDVEATVAEIHRLSEAGCOIVRVACPDERA 66  
QY AEAFLIKQOVNPLVADIHFDYRIALKVAEYGVDCILRNPNIGNERIRMVVDCARDK 128  
DB 67 ANLSAIKKKIHPLVADIHFDYRIALKAIDACVDYKIRINPNIGNRDRVEKVNNAKAK 126  
QY 129 NIPRIGVAGSLIEKQERKYEPTPQALLSEAMRHVDHLDRINPQFVSVKASDVFLA 188  
DB 127 NIPRIGVAGSLIEKQERKYEPTPQALLSEAMRHVDHLDRINPQFVSVKASDVFLA 186  
QY 189 VESYRLAKQIDQPHLGTTEAGARSGAVKASIGIGLLSEIGTTLVSLAADVEEL 248  
DB 187 IEAYDKASRAFNPLHLGITESTGTOPAGGKSAAGAILSLIGNTLVSLADVEEL 246  
QY 249 KVGFILKSLRIRSRGINFIACPTCSROEFVDYGVNVALEORLEDTITPMDSIICVYN 308  
DB 247 KVAREVLKSPGLSSNAAMLISCTGRIIDILRIANEVENYIAKEVIRKAVAGCAVN 306  
QY 309 GPEGALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD EARRIDVQ 368  
DB 307 GPEAREADIGTAGSVGEGLLFRHGKIIKRVPEALMIDELKKEIDILAEF-FVKKIDLE 365  
QY 369 QV 370  
DB 366 SL 367

RESULT 4  
ABP65322  
ID ABP65322 standard; Protein, 403 AA.  
XX  
XX ABP65322;  
XX  
XX 19-NOV-2002 (first entry)  
XX  
XX Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:66.  
DE Bifidobacterium longum NCC2705  
XX  
XX Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;  
KM anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;  
KM identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;  
XX rotavirus; food composition; pharmaceutical composition.  
XX  
XX Bifidobacterium longum.  
XX  
XX EP1227152-A1.  
XX  
XX 31-UTL-2002.  
XX  
XX 30-JAN-2001; 2001EP-0102050.  
XX  
XX 30-JAN-2001; 2001EP-0102050.  
XX  
XX (NEST ) SOC PROD NESTLE SA.  
XX  
XX MPI; 2002-668397/72.  
XX  
XX Novel polynucleotide comprising Bifidobacterium genome sequence useful  
PT as a probe or primer for detecting and/or identifying Bifidobacterium  
PT longum in a biological sample -  
XX  
XX  
XX Claim 3; SEQ ID 66; 80pp; English.

CC The present invention describes a polynucleotide (I) comprising a  
CC sequence of a Bifidobacterium genome selected from the nucleotide  
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at  
CC least 90% identity or which hybridises with the sequences given in  
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding  
CC a fusion protein, comprising a sequence selected from 1097 sequences

PT increasing isoprenoid content and identifying e.g. antimicrobial  
 PT agents, comprises using DNA sequences from bacteria or parasites  
 PS  
 XX Disclosure; Page 15-17; 36pp; German.

XX This invention describes a novel method for incorporating gcpe and yfegB  
 CC genes into viruses and cells for increasing isoprenoid content and  
 CC identifying e.g. antimicrobial agents, comprises using DNA sequences (I)  
 CC from the gcpe or yfegB genes of bacteria or parasites or DNA sequences  
 CC (II) which hybridize to the specified genes or encode a protein  
 CC with the same biological activity as those encoded by the genes. The  
 CC invention also describes (1) plant cells containing (I) or (II); (2)  
 CC transformed plant cells, and transgenic plants regenerated from them,  
 CC that contain (I) or (II); (3) determining the enzymatic activity of a  
 CC gcpe protein; or (4) screening compounds (A) that have antimicrobial,  
 CC antiparasitic or antiviral activity in humans or animals or antiviral,  
 CC antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)  
 CC are used: (i) to increase the isoprenoid levels in viruses and cells;  
 CC (ii) for determining the enzymatic activity of gcpe and yfegB proteins;  
 CC and (iii) to identify compounds that inhibit activity of gcpe, i.e.  
 CC potential antibacterial, antimycotic, antiparasitic or antiviral agents  
 CC for use in humans or animals, or antiviral, antiparasitic, fungicidal  
 CC or herbicidal agents for agriculture.

XX Sequence 372 AA:

Query Match 100.0%; Score 1866; DB 22; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-181;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNOAPIORRKRSTRIYGVNVPIDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60  
 DB 1 MHNOAPIORRKRSTRIYGVNVPIDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60  
 QY 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCIRINPGNIGNEERIRM 120  
 DB 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCIRINPGNIGNEERIRM 120  
 QY 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSARHVDHLDRINFQFRVSV 180  
 DB 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSARHVDHLDRINFQFRVSV 180  
 QY 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSATIGLLSISGIDTLRVSL 240  
 DB 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSATIGLLSISGIDTLRVSL 240  
 QY 241 AADPVEIKVGFILKSLRIRSGINFIACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300  
 DB 241 AADPVEIKVGFILKSLRIRSGINFIACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300  
 QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDLDNNMDIQLLEARIKASQOLD 360  
 DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDLDNNMDIQLLEARIKASQOLD 360  
 QY 361 EARRIDVOQVEK 372  
 DB 361 EARRIDVOQVEK 372

RESULT 2

ID AAE19653 standard; Protein; 372 AA.

XX AAE19653;

DT 31-MAY-2002 (first entry)

XX Escherichia coli GCPE protein.

KW gcpe gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;  
 KM transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;  
 KM food; feed source; transfection; single nucleotide polymorphism; SNP;  
 KM oxidative stress tolerance; UV tolerance; transformation; GCPE protein;

KW plant.

XX Escherichia coli.

XX WO200212478-A2.

XX 14-FEB-2002.

XX 06-AUG-2001; 2001WO-US24335.

XX 07-AUG-2000; 2000US-223483P.

XX (MONSANTO TECHNOLOGY LLC.

XX Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;

XX Valentin HE, Venkatesh TV, Venkatramesh M;

XX WPI; 2002-227151/28.

XX gcpe nucleic acid which is an essential gene of the methyl-D-erythritol

PT phosphate pathway, encoding a fully defined GCPE protein which is

PS useful for increasing levels of tocopherol substrates in plants

PS Claim 1; Page 144-145; 155pp; English.

XX The invention relates to gcpe nucleic acid molecule, an essential gene  
 CC of methyl-D-erythritol phosphate (MEP) pathway that encodes rice,  
 CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpe is useful  
 CC for producing a transgenic plant such as Brassica campestris, B. napus,  
 CC canola, castor bean, coconut, cotton, crambe, linseed, maize, sunflower,  
 CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,  
 CC or wheat with an increased isoprenoid (tocopherol) compound level. The  
 CC expression of GCPE protein in organisms increases the level of  
 CC tocopherol substrate such as isopentyl diphosphate and dimethylallyl  
 CC diposphate biosynthesis. Transgenic organisms overexpressing GCPE  
 CC protein can nutritionally enhance food and feed sources. Overexpression  
 CC of GCPE protein in transgenic plant may provide tolerance to stresses  
 CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV  
 CC tolerance, etc. gcpe may be used to obtain nucleic acid molecules from  
 CC the same species, and to obtain nucleic acid homologues. gcpe is also  
 CC used as or primers. The recombinant vectors are used in plant  
 CC transformation or transfection. gcpe an also act as markers capable of  
 CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).  
 CC gcpe is also used to determine the level or pattern of expression of  
 CC the protein. The present sequence is Escherichia coli GCPE protein.

XX Sequence 372 AA:

Query Match 100.0%; Score 1866; DB 23; Length 372;  
 Best Local Similarity 100.0%; Pred. No. 7.6e-181;  
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNOAPIORRKRSTRIYGVNVPIDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60  
 DB 1 MHNOAPIORRKRSTRIYGVNVPIDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60  
 QY 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCIRINPGNIGNEERIRM 120  
 DB 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCIRINPGNIGNEERIRM 120  
 QY 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSARHVDHLDRINFQFRVSV 180  
 DB 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSARHVDHLDRINFQFRVSV 180  
 QY 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSATIGLLSISGIDTLRVSL 240  
 DB 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSATIGLLSISGIDTLRVSL 240  
 QY 241 AADPVEIKVGFILKSLRIRSGINFIACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300  
 DB 241 AADPVEIKVGFILKSLRIRSGINFIACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using SW model

Run on: November 23, 2003, 15:54:02 ; Search time 51 Seconds  
(without alignments)  
1157.770 Million cell updates/sec

Title: US-09-921-992-50  
Perfect score: 1866  
Sequence: 1 MHNQAPIQRKRSTRIVGVN.....RAKASQLEARRIDVOQVEK 372

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_19Jun03:\*

- 1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*
- 2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*
- 3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*
- 4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*
- 5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*
- 6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*
- 7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*
- 8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*
- 9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*
- 10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*
- 11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*
- 12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*
- 13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*
- 14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*
- 15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*
- 16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*
- 17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*
- 18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*
- 19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*
- 20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*
- 21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*
- 22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*
- 23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*
- 24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	372	22	AB45692
2	1866	100.0	372	23	AB45692
3	868	46.5	368	23	AB45692
4	814.5	43.6	403	23	AB45692
5	794	42.6	359	19	AAW98298
6	785.5	42.1	415	22	AAU56212
7	761	40.8	378	22	AAU56212
8	751.5	40.3	387	22	AAU56212
9	751.5	40.3	387	24	AB57500

10	534	28.6	421	24	ABP78786
11	490	26.3	611	20	AAJ37164
12	470.5	25.2	621	20	AAJ34971
13	462	24.8	740	22	AAE19650
14	460	24.7	679	22	AA45693
15	446	23.9	752	21	AAW98883
16	445.5	23.9	603	23	AAE19652
17	445.5	23.9	666	23	AAE19651
18	428.5	23.0	169	21	AA451371
19	428.5	23.0	169	21	ABP32470
20	398	21.3	144	23	ABU52039
21	242	13.0	776	22	ABG25879
22	221	6.4	113	21	AAU08755
23	119	6.4	113	21	AAU08755
24	113	6.1	489	22	AAU96357
25	111.5	6.0	578	20	AAU89273
26	110	5.9	108	21	AAU08756
27	110	5.9	108	21	AAU08756
28	109	5.8	256	22	AAU53170
29	106.5	5.7	284	23	ABU53922
30	105.5	5.7	578	19	AAU82635
31	105.5	5.7	578	21	AAU78566
32	105.5	5.7	578	23	ABG93408
33	105.5	5.7	578	23	AAU73205
34	105	5.6	545	20	AAU26187
35	104	5.6	606	23	AB48094
36	102.5	5.5	542	23	ABU53701
37	102	5.5	546	23	ABP26321
38	102	5.5	546	23	ABP26321
39	101	5.4	338	23	AB48474
40	101	5.4	523	22	AAU61619
41	100	5.4	343	19	AAU71467
42	100	5.4	343	21	AAU99880
43	100	5.4	502	22	AAU99880
44	99	5.3	541	21	AAU91970
45	99	5.3	268	22	AAU34139

## ALIGNMENTS

RESULT 1	AA45692	standard; Protein; 372 AA.
AC	AA45692;	
AC	AA45692;	
DT	15-MAR-2001	(first entry)
DE	E. coli gcpe protein.	
XX		
KW	Isoprenoid; gcpe; yfG; antimicrobial; transgenic plant; agriculture; anti-mycotic; antiparasitic; antiviral; fungicidal; herbicidal.	
XX		
OS	Escherichia coli.	
XX		
PN	MO200072022-A1.	
PD	30-NOV-2000.	
XX		
PF	20-MAY-2000; 2000WO-EP04592.	
XX		
PR	21-MAY-1999; 99DE-1023567.	
PR	21-MAY-1999; 99DE-1023568.	
PA	(JOMA/) JOMA H.	
XX		
PI	Joma H;	
XX		
XX	WPI; 2001-025196/03.	
DR	N-PSDB; AAC82653.	
XX		
PT	Incorporating gcpe and yfG genes into viruses and cells, for	